

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:07:18 ; Search time 1381.67 Seconds
(without alignments)
-3065.678 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgttcacaaataccaata.....tcaggccaagtcccgtoaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :					GenEmbl.*				
					1: gb_bal.*	2: gb_ba2.*	3: gb_om.*	4: gb_ov.*	5: gb_pat.*
					6: gb_ph.*	7: gb_p11.*	8: gb_p12.*	9: gb_p1.*	10: gb_pr2.*
					11: gb_pr3.*	12: gb_ro.*	13: gb_sts.*	14: gb_sy.*	15: gb_un.*
					16: gb_vi.*	17: em_fun.*	18: em_hum1.*	19: em_hum2.*	20: em_in.*
					21: em_om.*	22: em_or.*	23: em_ov.*	24: em_pat.*	25: em_ph.*
					26: em_pl.*	27: em_ro.*	28: em_sts.*	29: em_sy.*	30: em_un.*
					31: em_vi.*	32: gb_htg1.*	33: gb_htg2.*	34: gb_in1.*	35: gb_in2.*
					36: em_bal.*	37: em_ba2.*	38: em_hum3.*	39: em_hum4.*	40: gb_pr4.*
					41: gb_htg3.*	42: gb_htg4.*	43: gb_htg5.*	44: gb_htg6.*	

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	365.6	26.2	1961	2	PAU32853	U32853 Pseudomonas
2	365.6	26.2	4587	2	PAU49151	U49151 Pseudomonas
3	274.8	19.7	2027	1	AVU30799	U30799 Azotobacter
4	212.4	15.2	1974	2	RMU31512	U31512 Rhizobium m
5	211.8	15.2	2009	1	BUU07352	U07352 Brucella ab
6	197	14.1	1905	1	BRUHTRAH	L09274 Brucella ab
7	196	14.1	2714	2	ECU32495	U32495 Escherichia
8	196	14.1	3156	2	ECU15661	U15661 Escherichia
9	196	14.1	10713	2	AE000402	AE000402 Escherich
10	196	14.1	110000	2	ECOUW67_1	Continuation (2 of
11	186.6	13.4	1191	2	PAU29172	U29172 Pseudomonas
12	178	12.8	10199	2	AE000125	AE000125 Escherich
13	178	12.8	82727	1	ECO82K	D26562 Escherichia
14	178	12.8	123171	1	ECU70214	U70214 Escherichia
15	176.4	12.6	1855	1	ECHTRA	X12457 Escherichia
16	176.4	12.6	1855	1	ECOHTRAA	M36536 E.coli htra
17	175.6	12.6	1980	1	STSPHSPG	X54548 S.typhimuri
18	175.6	12.6	1980	5	AR037634	AR037634 Sequence
19	175.2	12.6	1980	5	A18802	A18802 htra gene.
20	164.6	11.8	4151	1	RCFDXE	Y11304 R.capsulatu
21	164.4	11.8	1800	1	D78376	D78376 Yersinia en
22	164	11.8	110000	44	AC009220_1	Continuation (2 of
23	157.2	11.3	1613	1	BAU07351	U07351 Brucella ab
24	154.2	11.1	1791	5	AR000272	AR000272 Sequence
25	154.2	11.1	1791	5	I51141	I51141 Sequence 7
26	154.2	11.1	1791	5	I78681	I78681 Sequence 7
27	154.2	11.1	1889	1	ROCHTRA	L20127 Rochalimaea
28	149.6	10.7	11201	2	AE001732	AE001732 Thermotog
29	148.2	10.6	11881	2	AE001355	AE001355 Chlamydia
30	147.6	10.6	2778	1	BUJ13616	Y13616 Bradyrhizob
31	139.6	10.0	33773	41	AC008862	AC008862 Homo sapi
32	139.4	10.0	1509	2	AF018152	AF018152 Haemophil
33	139.4	10.0	2894	5	AR034125	AR034125 Sequence
34	139.4	10.0	2894	5	I19541	I19541 Sequence 1
35	139.4	10.0	2894	5	I60405	I60405 Sequence 1
36	139.4	10.0	2894	5	I64380	I64380 Sequence 1
37	138.8	9.9	1506	1	YEHTRA	X94153 Y. enterocol
38	138.4	9.9	10477	2	AE001474	AE001474 Helicobac
39	137.8	9.9	1314	2	AF018151	AF018151 Haemophil
40	137.8	9.9	12480	2	U32805	U32805 Haemophilus
41	136.2	9.8	1611	5	AR061048	AR061048 Sequence
42	136.2	9.8	1611	5	I70313	I70313 Sequence 1
43	136.2	9.8	1611	5	I89266	I89266 Sequence 1
44	133.4	9.6	13485	2	AE001678	AE001678 Chlamydia
45	131.6	9.4	42688	41	AC009092	AC009092 Homo sapi

ALIGNMENTS

RESULT 1
PAU32853 PAU32853 1961 bp DNA BCT 06-MAR-1996
LOCUS Pseudomonas aeruginosa mucC and mucD genes, complete cds.
DEFINITION U32853
ACCESSION U32853
VERSION U32853.1 GI:1184682
KEYWORDS

TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Mechanism of conversion to mucoidy in <i>Pseudomonas aeruginosa</i> infecting cystic fibrosis patients Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8377-8381 (1993) 93391358	-10_signal	sigmaE promoters of <i>E. coli</i> in both the -10 and -35 sequences" 411. .415 /note="P1 promoter" 421
	4 (bases 252 to 488) Martin,D.W., Schurr,M.J., Yu,H. and Deretic,V. Analysis of promoters controlled by the putative sigma factor AlgU regulating conversion to mucoidy in <i>Pseudomonas aeruginosa</i> : relationship to sigma E and stress response J. Bacteriol. 176 (21), 6688-6696 (1994) 95050230	mRNA	/note="transcription start site for promoter P1; 3' end is unknown" 474. .1055 /gene="algU" 474. .1055 /gene="algU" /codon_start=1 /transl_table=11 /function="conversion to mucoidy; required for algD transcription" /product="sigma factor" /protein_id="AAC43714.1" /db_xref="GI:1173502"
	5 (sites) Yu,H., Schurr,M.J. and Deretic,V. Functional equivalence of <i>Escherichia coli</i> sigma E and <i>Pseudomonas aeruginosa</i> AlgU: <i>E. coli</i> rpoE restores mucoidy and reduces sensitivity to reactive oxygen intermediates in algU mutants of <i>P. aeruginosa</i> J. Bacteriol. 177 (11), 3259-3268 (1995) 95286510	gene	/translation="MLTQEQDQLVERVQQRKAFDILLVLYKIQHKLIGLIVRFVHDA QEQDVAQAFIKAYRAGNFRGDSAFYTWLYRIATKNHLVARGRPDSDVTAE DAEFEGDHAKDIESPERAMLRDEIATVHTQIQLPEDLRLTALTTLREFEGLSYEDI ATVMQCCVGVVRSRIFRAREIDKALQPLREA" 1087. .1671 /gene="mucA" 1087. .1671 /gene="mucA"
	6 (bases 1 to 252) Schurr,M.J., Yu,H., Boucher,J.C., Hibler,N.S. and Deretic,V. Multiple promoters and induction by heat shock of the gene encoding the alternative sigma factor AlgU (sigma E) which controls mucoidy in cystic fibrosis isolates of <i>Pseudomonas aeruginosa</i> J. Bacteriol. 177 (19), 5670-5679 (1995) 96032406	CDS	/codon_start=1 /transl_table=11 /product="MucA" /protein_id="AAC43715.1" /db_xref="GI:1173503"
TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Two distinct loci affecting conversion to mucoidy in <i>Pseudomonas aeruginosa</i> in cystic fibrosis encode homologs of the serine protease HtrA J. Bacteriol. 178 (2), 511-523 (1996) 96134987	gene	/translation="MSREALQETLSAVNDNEADELELRVLAACGEDAEELRSTWRSYQ LARSVMHREPTLPKLDIAAAVSAALADEAPPKAEKGPWVMVGRVLAASVTLAVLAG VRLYNONDALPQMAOQGTTPQIALPOVKGPVAVLAGYSEEQAGAPQVITNSSSDTRWE QRPLVLRQHVQQSAVSGSTESALPYARAASLENR" 1680. .2630 /gene="mucB" 1680. .2630 /gene="mucB"
	8 (bases 1 to 4587) Boucher,J.C., Martin,D.W., Schurr,M.J., Deretic,V., Yu,H., Mudd,M. and Martinez-Salazar,J. Direct Submission Submitted (09-FEB-1996) John C. Boucher, Microbiology, U.T. HSC San Antonio, 7703 Floyd Curl Dr., San Antonio, TX 78284, USA On Mar 11, 1996 this sequence version replaced gi:1173501. Location/Qualifiers 1. .4587 /organism="Pseudomonas aeruginosa" /strain="PAO1" /db_xref="taxon:287" /map="67.5"	CDS	/codon_start=1 /transl_table=11 /product="MucB" /protein_id="AAC43716.1" /db_xref="GI:1220192"
	22 /note="transcription start site for promoter P5; 3' end is unknown"	gene	/translation="MRTSLLLGLSGMAVPATQAADSWLNRLAEADRONSFQOTF VYRNGSFSTHEIWHRVSDGAVRERLIQDGAQEVVVRDGTQCISGLQADQLADA QLWPKRFDPQSALSWYDLRLVGSVRVAGRPVAVLAVTPRQHRVGFELHLDRTGLP LKSILLNEKGOLLERFOFTOLNTGAAPAEODLOQAGECQVGPAKADSEKTYVNRSEW LPFGFTLITRFMRSPVTPDPVACLTGDLGRFSVFIEPLHGAMVGMDARSOLGPTVW VSKRLQDDGGQVMVTVVGEVPLGTAEVVALSIRPEAAQAQ" 2627. .3082 /gene="mucC" 2627. .3082 /gene="mucC"
	124 /note="transcription start site for promoter P4; 3' end is unknown"	CDS	/codon_start=1 /transl_table=11 /product="MucC" /protein_id="AAC43717.1" /db_xref="GI:1220193"
-35_signal -10_signal mRNA	204. .238 /note="P3 promoter" 204. .209 /note="P3 promoter" 226. .230 /note="P3 promoter" 235 /note="AlgU-dependent transcription start site for promoter P3; 3' end is unknown"	gene	/translation="MIEQGRVVAEPGAVVETVRRSTCCSCSANACGGQGLMORLG VGAGRVRVALSDLSLRVGDVNLGHEDLLLRASVLFPLFGFFVAALATRAGL VEPLIIVSGIAGLLAALLVRRHARRHADDPASPVVLRALISGPSDSA" 3122. .4546 /gene="mucD" 3122. .4546 /gene="mucD"
	389. .394 /note="P1 promoter" 389. .421 /note="P1 promoter, the first of two AlgU-dependent promoters; AlgU-dependent promoters strongly resemble the	CDS	/codon_start=1 /transl_table=11 /product="MucD" /protein_id="AAC43718.1" /db_xref="GI:1220194"
		gene	/translation="MHTLKRCAAMVALLSLAMTARAEPLDFTPLVEQASPAVYNI STQKLPDRAMARGQLSIDPLEGLPMPFRFLERSIPQVPRNPRQQREQSLSGSFI
		CDS	

ISNDGYLLTNHVVADADEILVRLSDRSEHKAKLIGADPSDVAVLKIENKNLPTLKL GDSNKLKGEWLAIGSPFGFDSHTAGIVSAKGRSLPNRESYVFFQTQDVAIINPNSG GPLNLQGEVVGINSQIFTRSGFGMLGSAIPIDVALNVADQLKKGKAGSRGWLGVVI QEVNKDLAESFGLNDVEDGPAAGKGLQGVQDVLISLNGOSINRGADLPHL VGNMKPGDKINDVIRNGRKSLSMAVGLSDDDDEETIASMGAPGARSNNRLGVTYVAD LTAQRKSLDIOGGVVIKEVQDGPAAVIGLRPGDVITHLDNKAVTSTKVPADVAKALP KNRSVMRLRQGRASFITFKLAE"										
BASE COUNT	811 a	1447 c	1515 g	814 t						
ORIGIN										
Query Match	26.2%	Score 365.6	DB 2	Length 4587						
Best Local Similarity	68.0%	Pred. No. 5e-65								
Matches 525	Conservative 0	Mismatches 244	Indels 3	Gaps 1						
QY	374	tcggttcgggttcctcatcagcaaaacggctacatctctgaccataaccacgctcgttg	433							
Db	3411	TGGGTTCCGGCTTCATCTCTCAACGACGGCTACATCCTCACCAACAATCACGTCGTGG	3470							
QY	434	cgggtatgggcagtatcaaaagtctctgctcaacgacagcgcggaatataccgccaactca	493							
Db	3471	CCGATGCCGACGAGATCCTGCTGGCCCTGTCCGACCGTAGCGAGCACAAAGGCCAAGTTGA	3530							
QY	494	tcggttcgagtgatccaatccgagtgctgcctctctgaaatcgcagcgaacgaagctac	553							
Db	3531	TCGGCGGGACCCGGCAGGAGCTGGCGGTGTGAAGATCGAGG---CGAAGAACCTGC	3587							
QY	554	ccgtcgtcaaaatcggcaatcccaaaatttgaacccggcggaatgggtcgtgccatcg	613							
Db	3588	CGACCTGAAACTGGCGGATTCGAACAAGCTGAAAGTGGCGAATGGGTCTCGGCCATCG	3647							
QY	614	ggcgcccttcggttttgacaaacagcgtgacccgcgcatcgtgtccgcgcaaaaggcgaa	673							
Db	3648	GTTCCGCGTTCGGCTTCGATFCATCTCGGTCAACCGCGGTATCGTCAGTGCCAAAGGTCGTA	3707							
QY	674	gctgcgccaacgaagatcacacccttcatacaaacgacgcttgccatcaatccgggca	733							
Db	3708	GCCTGCGGAACGAGAGCTACGTACCCTTCATCCAGACCGGAGTGGCGATCAACCCGGGTA	3767							
QY	734	attccggcgccgctgttctcaacttaaaaggacaggtcgtcgccatcaattcgcaaatat	793							
Db	3768	ACTCCGGCGTCCGCTGCTGAACCTGCGAGGCGAAGTGTGCGCATCAACTCGCAGATCT	3827							
QY	794	acagcgcgagcgcggtattcatatgggcatctcctttggcatcccgattgagcttgccatga	853							
Db	3828	TCACCCGTTCCGGCGGCTTCATGGCGCTGCTCTCGCCATTCGATCGATGTCGCGTGA	3887							
QY	854	atgtcgccgaacagctgaaacacaccgcggaagtcacacgcggacacactggcgctgatta	913							
Db	3888	ACGTCCCGACCAAGTTGAAGAAGCCGCAAGGTCACTCGCGGTGCTGGCGGTGGTGA	3947							
QY	914	ttcgaagaagtatacctcagtttggcacagtcgttctgctctggataaagcagcgcgcat	973							
Db	3948	TCCAGGAAGTGAACAAGGATCTCGCGAGTCTTCGCGCTTCGACAAAGCCGTCGCGCGCGC	4007							
QY	974	tgattcccaaaatccttcccgagccccgcagaaagtcgcgccctcagcgggcgcaaca	1033							
Db	4008	TGTTGCGCAACTGGTGGGAAGACGGCCCGCGCCCAAGGGTGGCGCTTCAGGTGGCGCATG	4067							
QY	1034	tgctcctcagcctcgacggcgggagaaatacgtttcttcgcgacacttcccgatcgctg	1093							
Db	4068	TGATCTCAGCCTGAACGGCCATCGATCAACGAGTCCGCGACCTCGCGCACCTGGTGG	4127							
QY	1094	ggccattacgcgggaagaagtcagcctcgcgctatggcgcaaaaggcga 1145								
Db	4128	GCAACATGAAGCGCGGCGACAAGATCAACCTGGACGTGATTCGCAACGGCCA 4179								
RESULT	3									
AVU030799	2027 bp	DNA	BCT	31-MAY-1996						
LOCUS	Azotobacter vinelandii MucC (mucC) and MucD (mucD) genes, complete									
DEFINITION	cds.									

ACCESSION	U30799										
VERSION	U30799.1	GI:1345102									
KEYWORDS	Azotobacter vinelandii.										
SOURCE	Azotobacter vinelandii										
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae; Azotobacter.										
REFERENCE	1 (bases 1 to 2027)										
AUTHORS	Martinez-Salazar,J.M., Moreno,S., Najera,R., Boucher,J.C., Espin,G., Soberon-Chavez,G. and Deretic,V.										
TITLE	Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA, MucB, MucC, and MucD in Azotobacter vinelandii and evaluation of their roles in alginate biosynthesis										
JOURNAL	J. Bacteriol. 178 (7), 1800-1808 (1996)										
MEDLINE	96178940										
REFERENCE	2 (bases 1 to 2027)										
AUTHORS	Martinez-Salazar,J.M., Moreno,S., Najera,R., Boucher,C., Espin,G., Soberon-Chavez,G. and Deretic,V.										
TITLE	Direct Submission										
JOURNAL	Submitted (30-JUN-1995) Jaime M. Martinez-Salazar, Microbiology, UTHSCSA, 7703 Floyd Curl Drive, San Antonio, TX 78284, USA										
FEATURES	Location/Qualifiers										
source	1. .2027										
	/organism="Azotobacter vinelandii"										
	/strain="UW 136"										
gene	/db_xref="taxon:354"										
	1. .471										
	/gene="mucC"										
CDS	1. .471										
	/gene="mucC"										
	/note="similar to Photobacterium sp. ORF4 product, encoded by GenBank Accession Number L41667"										
	/codon_start=1										
	/transl_table=11										
	/product="MucC"										
	/protein_id="AAB01512.1"										
	/db_xref="GI:1345103"										
	/translation="MIEPGRVATIEREAVNIEALDEAACPGCTASVCGGRWSFDWLE TPRRRRIRALSVNPLRVGDAVVYGIREDLFLRSTLAIYLLPLGLFVCTIAIOQLGFGSGLMIPAGLFGVSVWAVRRISRRQNMNDPTLQPVVLRALPGASEGVQRAAF"										
	483. .1904										
CDS	/note="similar to Brucella abortus HtrA protein encoded by GenBank Accession Number U07352"										
	/codon_start=1										
	/transl_table=11										
	/function="serine protease"										
	/product="MucD"										
	/protein_id="AAB01513.1"										
	/db_xref="GI:1345104"										
	/translation="MSKVGFKSFASVLAGALLQGSFLVQAQLPEFTSLVEASPAVY NISTROKLPDRSTVQGLPDLEGLPLREFLESIPOLPPTPDNGRQREAHSLGSGFI ISPDYVLTNNHVVADAEIIVRLSDRSELAELVGDPLTRCSFEVQGGFSPHSQGTGYRTIESRGMGSHRRPFRSFRDCGHCHGKPAERELVPIOTDVAINPNSGGLPFLDGRVIGINSQIFTRSGGFMGLSFAIPIEVAMGVADOLKATGKPTVMASDLPHL QEYNKDIAESFGLDRPAGALVAQVLEDGPADKGGLOQGDVILSDGHPITVMSADLPHLVGGILKPGAAANLEVVDRGNRIIAITVGALPEEINGVQPSIAGTQSSNRNLGVTVELTAEQKSLDLKGWVIRELVNGNPAALIGLRPDVVVTHLNNQPIDSAKTFAEVAGALPKGRVSMRVLQRGRASFITFKLAE"										
	GRVSMRVLQRGRASFITFKLAE"										
BASE COUNT	366 a	529 c	658 g	474 t							
ORIGIN											
Query Match	19.7%; Score 274.8; DB 1; Length 2027;										
Best Local Similarity	58.9%; Pred. No. 1.5e-46;										
Matches 527; Conservative	0; Mismatches 362; Indels 5; Gaps 3;										
QY	295	ttctacgaatttttcaaacgcctcgtccgacaatcgccgaaatcccccaagaagca	354								
Db	693	TTCAGGAGTTCTTGGACGCGAGCATTCGCAACTTCCGCGTACTCCGGATAACGGCGG	752								
QY	355	gatgacgcgcatgaaattcggttcgggtctcatcatcagcaaaaacggctacatcctg	414								
Db	753	CAGCGTAGGCGCACTCCCTCGGCTCGGGTTTCATCATTTCTCCAGATGGCTATGTCTA	812								


```

DEFINITION Brucella abortus immunoreactive stress response protein gene,
complete cds.
ACCESSION L09274
VERSION L09274.1 GI:144117
KEYWORDS homologue; immunoreactive stress response protein.
SOURCE Brucella abortus (strain 2308) (library: Sau3A/BamHI library of M.Roop) DNA.
ORGANISM Brucella abortus
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Brucella.
REFERENCE 1 (bases 1 to 1905)
AUTHORS Roop,R.M. II., Fletcher,T.W., Sriranganathan,N.M., Boyle,S.M. and Schurig,G.G.
TITLE Identification of an immunoreactive Brucella abortus HtrA stress response protein homolog
JOURNAL Infect. Immun. 62 (3), 1000-1007 (1994)
MEDLINE 94156447
FEATURES
    Location/Qualifiers
        1..1905
            /organism="Brucella abortus"
            /strain="2308"
            /db_xref="taxon:235"
            /tissue_lib="Sau3A/BamHI library of M.Roop"
            183..213
                /note="This region of nucleotide sequence shows a high level of homology with the sigma type E promoters previously described for the HtrA genes of E. coli and S. typhimurium and the rpoH gene of E. coli; putative"
                /function="sigma E type promoter"
                322..326
                333..407
                333..1835
                    /note="This 45 kD protein precursor is an immunoreactive protein, the deduced amino acid sequence shows 40.5% and 43.2% identity with the Escherichia coli and Salmonella typhimurium HtrA (DegP) stress respon"
                    /codon_start=1
                    /transl_table=11
                    /evidence=experimental
                    /product="immunoreactive stress response protein"
                    /protein_id="AAA53693.1"
                    /db_xref="GI:144118"
                    /translation="MSRARISNVRKGVAAVALSASLAGAFVVTGPLGALNEARAAVH VTPPGFADLVEKVRPAVSVRVKVKQVETSNRGPFQFFGPPGDLDPGLPKRFF RDEGPEGRSDSRNRKRPGRKHPERPAQGGFVISEDYVYVNNHVVSDGDAY TVLDDGTDLAKLIGADPTDLAVLKINAPKRFVYSPLEATIRCDVGDVAVGNP FGLGGTVTSIGVARGADIGAGPYDDFIQIDAANKNGSGGPAFLDSGVIGINTAIF SPGGTVGTAFAIPSTAKOVVDQLIKKGVBERGWIGVOIQPVTKDIAASLGLAEKG AIVASPDQGPAAKAGIKAGDVITAVNGETVODPDRLARKVANIAPGEKAAALTWFEK KAEINVTIAAMPNDKSGSQSNDNGGGETLDSYGLTVVPSEDKGVVYTDVDDP SDAADRISRGDVIYSVNNOTVKTAGDKINKAITAEKSRNGAAAIASQ"
                    408..1832
                    /evidence=experimental
                    /product="immunoreactive stress response protein"
mat_peptide
BASE COUNT 393 a 558 c 564 g 390 t
ORIGIN
Query Match 14.1%; Score 197; DB 1; Length 1905;
Best Local Similarity 54.6%; Pred. No. 1e-30;
Matches 436; Conservative 0; Mismatches 360; Indels 3; Gaps 2;

Qy 376 gtttcgggttcattcgcgaacaaacggctacatcctgaccataccacgtggtgccc 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 GGATCGGGCTCGTCATTCCGAAGATGGCTATGCTGTACCAACACCATGCTGCTTCC 797
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 436 ggtatggcagatcaaatcctcctcaacgacgaagcgccgaataaccacactcatc 495
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 GATGGCGATGCTATACCGTCGTGACGACCGACCGACCGACCTGTATGCCAAGCTGATT 857
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 496 ggttcggtatcccaatccgattgctgcttctgaaataccgacgaacg--gaagagctac 553
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 GGTCTGATCGCGCACGGACCTGGCCGTGTTGAAATCAACGCTCCGGAAGCGCAAGTTC 917

```

```

Qy 554 ccgtgctcaaatgggcaatcccaaaaatttgaaac-cggggaatgggtgctgctccatc 612
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 918 GTCTATTTCGCCTTTGGCGACGATATAAGGTGCGACGTCGGTGATTGGTTTTCGCAGTT 977
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 613 ggcggcccttcggtttgacacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 672
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 978 GGCAATCCGTTTCGGCTTGGTGGCACGGTGACTTCCGTAICGTTTTCGCCCGCTGGAGCC 1037
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 673 agcctggcccaacgaagctacacacccttcccaaaacgcgctgttgccatcaatccgggc 732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1038 GACATCGCGGACGACCCCTATGACGATTTCATCCAGATTGATCCCGCTGAACAAGGCG 1097
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 733 aattccggcgcccgctgttcaacttaaaagacagggctgctgacatacaatcgcgaata 792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1098 AATTCGGTGTGTCGGCTTTCGACCTTTCCGGCGAAGTCATCGGCATCAACACCGGCATC 1157
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 793 tacagccgagcgcggtattcggcatctcttggcatcccgatcccgattgacgttgcctg 852
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1158 TTCTCGCCATCTGGCGGACGGTGGCATGTGTTTTCGCAATTCGCTCTCCACCGCCAG 1217
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 853 aatgtcgcgcgacagctgaaaaacacccgcaaaagtcacacgcggagcaacatggcgctgatt 912
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1218 CAGGTCTGCGACGAGCTCATCAAGAGGTTCCGTCGAGCGCGGTGATCGCGCTCCAG 1277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 913 attcaggaagtatcctacagctgttgcaacagtgctgctgctgataaaacccagcgcgca 972
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1278 ATACAGCTGTGACCAAGGATATCGCGCTTCGCTCGTCTTCCGAGGAAAGAGGTGCG 1337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 973 ttgattgcaaaatccttcccgcgaccccgacgaacgctgcccggcctgcagcgcgcgac 1032
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1338 ATCGTGGCTTCGCGGAGGATGATGCCCGGCTGCCAAGCTGGCATCAAGCGGCGGAT 1397
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1033 atcgtcctcagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1092
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1398 GTGATTACGGCTGTCAATGGCAACGGTTTCAGGACCGCGTACCTGGCGCGCAAGGTT 1457
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1093 ggcgcctacccggggaagaagtcacccctcggtgctgacgcaaaaggcggaagaatc 1152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1458 GCAATATCGCCCTGTGTAAGGCGGCGCTGACCGCTGCGCGCAAGAAAGCGCGAG 1517
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1153 acaatcaagcgaagctcg 1171
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1518 GAATCAACGTCAGATTG 1536
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS ECU32495 2714 bp DNA BCT 25-APR-1996
DEFINITION Escherichia coli putative serine protease (degQ and degS) genes, complete cds.
ACCESSION U32495
VERSION U32495.1 GI:984378
KEYWORDS .
SOURCE .
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2714)
AUTHORS Waller,P.R. and Sauer,R.T.
TITLE Characterization of degQ and degS, Escherichia coli genes encoding homologs of the DegP protease
JOURNAL J. Bacteriol. 178 (4), 1146-1153 (1996)
MEDLINE 96165272
REFERENCE 2 (bases 1 to 2714)
AUTHORS Waller,P.R.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1995) Patrick R.H. Waller, Biology, Massachusetts Institute of Technology, Room 68-565, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
FEATURES
    source
        1..2714
            /organism="Escherichia coli"

```

```
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/clone_lib="Kohara phage"
56..1577
/gene="degQ"
56..61
/gene="degQ"
79..84
/gene="degQ"
91
/misc_feature
91
/note="degQ"
/note="primer extension identifies this as the
transcription start site (+1 position)."
126..138
/gene="degQ"
136..216
/gene="degQ"
/evidence=experimental
136..1503
/gene="degQ"
/note="serine protease in vitro"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DegQ"
/protein_id="AAC44005.1"
/db_xref="GI:984379"
/translation="MKQTQLLSALSVGLTISASFOAVASIPGOVADAPLPSLAP
MLEKVLPAVSVRVEGTASOGKIPFEKFFGDDLPDQPAQPEFGSGSVIINASKG
XYLTNNHVINQAQKISIQLNDRREFDAKLIGSDQSDIALLOTONPKLQIAIADSD
KLKRGFAVAGVPFGQGTATSGIVSALGRSGNLNLEGNFIOTDASINRNGSGGAL
LNLNGELIGNTAILAPGGSVGIGFAIPSNMARTLAQQLIDFGEIRGLGKIGTEM
SADIAKAFNLVDQGFVSEVLPGSGSAGAKVAGDIITSLNGKPLNSFAELRSRIAT
TEPTKVKLLNKRKPLEVEVTLDTSTSSASAEITTPALEGATLSDGOLKDGKGI
KIDVVKGSPAAQGLQKDDVIICVNRDRVNSIAEMRKVLAAPAILALQIVRGNESI
YLLMR"
217..1500
/gene="degQ"
/note="serine protease in vitro"
/evidence=experimental
/product="DegQ"
1508..1547
/gene="degQ"
1508..1577
/gene="degQ"
/note="consistent with degQ transcript size defined by
Northern analysis; Rho independent: stem loops followed by
a series of t residues"
1532..1537
/gene="degS"
1532..2709
/gene="degS"
1550..1571
/gene="degS"
1554..1559
/gene="degS"
1566
/gene="degS"
/note="primer extension identifies this as the
transcription start site (+1 position)."
1579..1595
/gene="degS"
1593..2660
/gene="degS"
/note="putative serine protease"
/codon_start=1
/transl_table=11
/product="DegS"
/protein_id="AAC44006.1"
/db_xref="GI:984380"
/translation="MFVKLLRSVAIGLVGAILLVAMPFSRLSNPLSTPQFDSTDETP
ASYNLAVRRAPAVVNYNRGLNTNSHNQLEIRTLGSGVIMDQRGYIITNKHVINDAD
```

```
stem_loop
terminator

BASE COUNT 659 a 692 c 710 g 653 t
ORIGIN

Query Match 14.1%; Score 196; DB 2; Length 2714;
Best Local Similarity 53.6%; Pred No. 1.7e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 382 ggcttcacatcagcaaaacggctacatcctgaccacaataccaccgctgttgcggtatg 441
Db 418 gTCATCATCAACGCCAGTAAAGGCTATGTCTGACCAACAACCATGTGATTAAATCAGGCA 477
Qy 442 ggcagtatcaaaagtctgctcaacgacgacgaataataccgccaacatcaggttcg 501
Db 478 CAGAAAATCAGTATTTCAGTCAATGATGGCGCGAGTGTGATGCAAAACATGATTGGTAGC 537
Qy 502 ggtgccatcagtcgtccctctgaaatcgacgcaacggaagagtcaccgctgcgc 561
Db 538 GATGACGAGCGATATCCCTGTGTACAAATCAAAACCCGAGCAAAATTAACGCAAAATC 597
Qy 562 aaaaatcgggcaatccccaaaattgaaacccggcggaatgggtcgtccatcggcgcgccc 621
Db 598 GCTATTGCCGACTCCGATAAATTCGCGCTCGGTGATTTTGGCCGTAGCGTCAACCCA 657
Qy 622 ttgcgctttgacaacagcgtgaccgcgcatcgttcgcgcgaagggcagaagcctgcgc 681
Db 658 TTTGGCCTTTGGCAAAACCGCCACCTCTGGCATTGTTTCCGCAATTAGGCCGCGAGGTTG 717
Qy 682 aa---cgaaagctacacaccccttcacaaacgacgcttgcacatcaatcgggcaattcc 738
Db 718 AATCTTGAAGGTCGTGAAAAACCTTTATCCAGACAGATGCTTCCATTAAACCGCGTAACCC 777
Qy 739 ggcggccgcgtgttcaacttaaaaggacaggtcgtcggcatcaattcgcacaataacacgc 798
Db 778 GCGCGTGCACTATTAAACCTTAACGGTGAGTTAATTTGGCATCAACACTGCAATCCTTGCG 837
Qy 799 cgcagcggcggtattcatgggcatctccttggccatcccgattgacgttgcacgaatgctc 858
Db 838 CCTGGCGCGGAGCGTCGGGATTGGATTGGCATCCCGAGTAATATGGCGCGAAACACTG 897
Qy 859 gccgaacagctgaaaaaacacccggcaaaagtcacacgcggaacactggcggtgattattcac 918
Db 898 GCGCAGCAGCTTATCGACTTTGGTGAATCAAAACGGGTTTGTGTAGGCATCAAAAGGCACC 957
Qy 919 gaagtatctcaggtttggcacagtcgttcggtctggataaaagccagcgcgcatgatt 978
Db 958 GAGATGAGTGCGGATATCGCAAAAGCCTTCAACCTTTGACGTGCGAGCGTGGCGGTTGTC 1017
Qy 979 gccaaaatccttcccggaagccccgcagaaactgacgagcctgcagcggcgagacatcgtc 1038
Db 1018 AGCGAAGTGTTCGCGAGTTCTGGCTCGGCAAAAGCGGGCGTCAAAAGCGCGGATATTATT 1077
Qy 1039 ctccagcctcagcggcgagaaataagttcttcggcgagaccttccctcactcgttcggcgcc 1098
Db 1078 ACCAGCCTCAACGGCAAAACCGCTGAATAGCTTTGCTGAGTTGGCTCTCGTATCGCGACC 1137
Qy 1099 attacgcggggaaaaagaagtcagcctcggcgtatgcgcgaagggcgagaataatcaaatc 1158
Db 1138 ACCGAGCGGGGACCAAAAGTAGGCTTGGCTCTGCTGCTAACGCAACCAACCTGGAAGTA 1197
Qy 1159 aaagccaagctgggcaacgcccgc 1182
```


Db 1198 GAAGTGACGCTCGATACCAGCAC 1221

RESULT 8

ECU15661 3156 bp DNA BCT 25-APR-1996
LOCUS Escherichia coli serine protease (hhoA and hhoB) genes, complete
DEFINITION cds, and malate dehydrogenase (mdh) gene, partial cds.
ACCESSION U15661
VERSION U15661.1 GI:558911
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 3156)
AUTHORS Bass,S., Gu,Q. and Christen,A.
TITLE Multicopy suppressors of prc mutant Escherichia coli include two
JOURNAL HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated RipA
MEDLINE J. Bacteriol. 178 (4), 1154-1161 (1996)
REFERENCE 2 (bases 1 to 3156)
AUTHORS Bass,S.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1994) Steve Bass, Cell Genetics, Genentech Inc.,
460 Point San Bruno Blvd, South San Francisco, CA 94080-4990, USA
FEATURES Location/Qualifiers

1..3156
/organism="Escherichia coli"
/strain="W3110"
/db_xref="taxon:562"
/map="70 minutes"
454..459
477..482
/note="putative promoter"
534..614
/gene="hhoA"
534..1901
/gene="hhoA"
534..1901
/gene="hhoA"
/note="putative serine protease"
/codon_start=1
/transl_table=11
/product="hhoA"
/protein_id="AAC43992.1"
/db_xref="GI:558912"
/translation="MKQTOLLALSVGLTLSASTQAVASIPGVADQAPLPSLAP
MLEKLVPAVSVRVEGTASQKQIPEEFKFFGDDLPDQPAQPEGGLSGVINASKG
YVLTNNHVIQAQKISIQINDREFDAKLIGSDSDIALLIQNPBKLTQIAADSD
KLRYGDFAVAGNPFGLQGTATSGIVSALRSLNLEFNFIQTASINRNGSGAL
LNLNGELIGINTALAPGGSGVIGFAPSNMARTLAQLIDFGEIKRGLIGKTEM
SADIAKFNLDVORGFVSEVLPGSGAKAGVAGDIITSLNGKPLNSFABELRSRIAT
TEPTKVLGLLRNKGPLEVEYTLDTSTSSASAEMLTPALEGATLSDGOLKDGKGI
KIDVEVGSPPAQAGLQKDDVLIIGVNRDVRNSIAEMRKVLAAKPAIALLQIVRNEST
YLKMR"
615..1898
/gene="hhoA"
/product="hhoA"
1906..1945
1991..3058
/gene="hhoB"
1991..3058
/gene="hhoB"
/note="putative serine protease"
/codon_start=1
/transl_table=11
/product="hhoB"
/protein_id="AAC43993.1"
/db_xref="GI:558913"
/translation="MFVKLLRSVATGLIVGAILLVAMPSLRSLNPLSTPQFDSTDETP
ASYNLAIVRAAPAVVNVYINRGLNTNSHNLQELRTLGSGVIMDQRGYIITNKHVINAD

-35_signal

-10_signal

sig_peptide

gene

CDS

mat_peptide

stem_loop

gene

CDS

terminator

gene

CDS

BASE COUNT 768 a 807 c 828 g 753 t
ORIGIN

Query Match 14.1%; Score 196; DB 2; Length 3156;
Best Local Similarity 53.6%; Pred. No. 1.7e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 382 ggcttcatacagcaaaacggctacatcctgaccaataccacgctcgtgcggtatg 441
Db 816 GTCATCATCAACGCCGCTAAAGGCTATGCTGCTACCAACACCATGTGATTAACAGGCA 875
QY 442 ggcagtatcaaaagtcctgctcaacgcaacgcaataataccgcaaaactcatcggttcg 501
Db 876 CAGAAAATCAGATTTCAGCTCAATGATGGCGCGAGTTTGATGCAAACTGATTGGTAGC 935
QY 502 gatgtccaatccgatgtgcocctctctgaaaatcgacacgcaagagagctaccgctcgc 561
Db 936 GATGACGACGCGATATCGCCTGTTACAAATTCAAAACCCGAGCAAAATTAACGCAATC 995
QY 562 aaatcggaatcccaaaatttgaaacccggcgcaatgggctgcgtgccatcgcgcgcc 621
Db 996 GCTATTGCCGACTCCGATAAATTGCGGCTCGGTGATTTTCCGTAGCGGTGCGTAACCCA 1055
QY 622 ttcggctttgacacagcgtgacgcggtgcgtgtgcgcacaaaggcagaagcgtgcgc 681
Db 1056 TTTGGCCTTGGGCAACCGCCACCTCTGGCATTTGTTCCGCATTAGCGCGAGCGGGTTG 1115
QY 682 aa---cgaaagctacacaccccttcataccaaaccacgctgctgcatacaatccggcaattcc 738
Db 1116 AATCTTGAAGTCTGGAAAACCTTTATCCAGACAGATGCTTCCATTAAACCGCGTAACGCC 1175
QY 739 ggcggccgctgtcaacctataaaggacagctgcgcgcatacaattcgcaaatatacacgc 798
Db 1176 GCGGTGCACATATTAAACCTTAACGGTGAGTAAATTGGCATCAACACACATCTCTGGC 1235
QY 799 cgcagcgcggtattcatgggcatctcctttgccatcccgattgacgttgccatgaatgc 858
Db 1236 CCTGGCGCGGAGCGTGGGATTTGGATTTGCCATCCCGAGTAATATGCGCGGAACACTG 1295
QY 859 gccgaacagctgaaaaacacccggcaagtcacacgcgacacgtggcggtgattattcaag 918
Db 1296 GCGCAGCAGCTTATCGACTTTGGTGAATAACAACCGCGTTTGTAGGCATCAAGGCGACC 1355
QY 919 gaagtatcctacggtttggcacagctcgttcgtctggataaaagccagcgcgcatgatt 978
Db 1356 GAGATGAGTGGCGATATCGCAAAAGCCTTCAACCTTGAGGTGACGGTGGCGCGTTGTTC 1415
QY 979 gcaaaatcctcccgcgagccccgcgacacgtgcggcctgcagcgcgacatcgctc 1038
Db 1416 ACGGAAGTGTGCCAGGTCTTGGCTCGGAAAAGCGGCGTCAAAAGCGCGCATATTATT 1475
QY 1039 ctacgacctcgacgcgagagaaatacgtttctccgcgacacctcccgtaatggtcgcgcc 1098
Db 1476 ACCAGCCTCAACGCGCAAAACCGCTGAATAGCTTTGCTGAGTTGCTGCTCTCGTATCGGACC 1535


```
/db_xref="GI:1789622"
/db_xref="PID:g1789622"
/translation="MEGLIISVILITMGLYLKKNYKPOVLAAGIFLMGCVWLGF
CGVDPTRSSGLYVDIINEILRMSNRAGUGLSIMVGGFARTMERIGASRAMVSL
LSRPUKLRSYPIILSATYVIGQIMAOFTTSASGLGMLMLVTLFPTLVLSGVSRLSA
AVIATLRSIEGILNFIFAQVAGMKIATFFHYQLFVASCVISIASHFFVORA
FDKKDNHQAQKALNDQVPLYYAILPVNPLILMLGSLFLAHVGLMQSELHLVVV
MLLSLTVMFEFFPKHNLRETMDVDQAFEDGMGTQFANVTVLVVAGETFAKGLTTIG
TVDAVIRGAHSGILGIGVIMIALYALCAIVMSGNAPMSPASLIINIAAGLHP
AVVIMPHFATTLARASPIITAVVVVTSIAGVSPFAVVKRTAIPMAVGFVYVNIAT
ITLF"
complement(3666. .4163)
/gene="sspb"
/note="b3228"
/note="(3666. .4163)
/gene="sspb"
/complement(3666. .4163)
/function="regulator; Global regulatory functions"
/note="f165; 100 pct identical amino acid sequence and
equal length to SSPB_ECOLI SW: P25663"
/codon_start=1
/transl_table=11
/product="stringent starvation protein B"
/protein_id="AAC76260.1"
/db_xref="GI:1789623"
/db_xref="PID:g1789623"
/translation="NDLSQLTTPRRPYLLRAFYEWLLDNQLTPLHLYVDVTLPGVQVPM
YARGQIVLNAPRAVGNLEANDVRFNAREGGIPQVSVPLAAVLAIYARENGAGT
MFEPFAAYDEDTISMNDEASADNETVMSVIDGDKPDHDDTHPDDPEPPQPPRGRPA
LRVK"
complement(4166. .4194)
/note="factor Sigma70; predicted +1 start at 3374407"
/complement(4169. .4807)
/gene="sspa"
/note="b3229"
/complement(4169. .4807)
/gene="sspa"
/function="regulator; Global regulatory functions"
/note="f212; 100 pct identical amino acid sequence and
equal length to SSPA_ECOLI SW: P05838"
/codon_start=1
/transl_table=11
/product="regulator of transcription; stringent starvation
protein A"
/protein_id="AAC76261.1"
/db_xref="GI:1789624"
/db_xref="PID:g1789624"
/translation="MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSEIEHVEKD
NPQDLIDLNPNQSVPTLVDELTLWESRIIMEYLDERPFPPLMPVYPVARGESRLY
MHRTEKDWTLWNTIINGSASADAARKOLRELLAIAPVFGKPYFLSDPSLVDCY
LAPLWRLPOLGIESGPGAKELKGYMTRVFRSDSLASLTAEEREMRLGRS"
complement(4944. .4972)
/note="No predicted promoter"
/bound_moiety="Galr predicted site"
/complement(5202. .5594)
/gene="rpsl"
/note="b3230"
/complement(5202. .5594)
/gene="rpsl"
/function="structural component; Ribosomal proteins -
synthesis, modification"
/note="f130; CG Site No. 17596; 100 pct identical
aminoacid sequence and equal length to RS9_ECOLI SW:
P02363"
/codon_start=1
/transl_table=11
```

```
Query Match 14.1%; Score 196; DB 2; Length 10713;
Best Local Similarity 53.6%; Pred. No. 1.6e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 382 ggcttcacacgcaaaacggctacatctgaccaataaccacgcgtgtcgcggtatg 441
- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8412 gtcacatcaacgccagtaaggctatgtctgaccaacacacccatcgattatcaggca 8471
```

```
Qy 442 ggcaagtatcaaaagtcctgtcaacgacaaagcgcaataataccgccaaactccatcggtcg 501
Db 8472 CAGAAAATCAGATTTCAGCTCAATGATGGCGCGAGTTTGATGCAAAACTGATTGCTAGC 8531
Qy 502 gatcccaatccgatgtcgccctctgaaatcgacgcaacggaagagctacccgcgtc 561
Db 8532 GATCACCAGAGCGATATCGCCCTGTTACAAATTCAAAACCCGAGCAAAATTACGCAAAATC 8591
Qy 562 aaatcggcaatccccaaaaatttgaacccggcggaatgggtcgctgceatcgcgcgccc 621
Db 8592 GCTATTGCCGACTCCGATAAATTGCGCTCGGTGATTTTCCGTAGCGTCGTAAACCCA 8651
Qy 622 ttggcgtttgacaacacggtgacccggtcgctgtgtcccaagagcagaagcctcgccc 681
Db 8652 TTTGGCCTTGGGCAACCCGCCACCTCTGGCATTTGTTCCGATTTAGGCGCGAGGGTTG 8711
Qy 682 aa----gaaagctacacaccccttccaaaccgagcttgccatcaatccgggcaattcc 738
Db 8712 AATCTGAAGGCTCGGAAACTTTATCCAGACAGATGCTTCCATTACCGCGGTAACTCC 8771
Qy 739 ggccgcccgtcttcaactaaaaaggacaggtcgctcggcatacaattcgcaaatatcacgc 798
Db 8772 GCGGCTGCACCTATTAAACCTTAACGGTGAGTTAATTGGCATCAACACTGCAATCCTTGC 8831
Qy 799 cgacgcccggattcatgggcatctcctttgccatcccgattgacgttgccatgaatgc 858
Db 8832 CCTGCGCGCGGAGCGTCGGGATTGGATTTCGCATCCCCAGTAATATGCGCGCAACACTG 8891
Qy 859 gccgaacagctgaaaaaacacccgcgcaagtcacacgcgacaaactggcggtgatttcag 918
Db 8892 GCGCAGCAGCTTATCGACTTTTGGTGAATCAACCGCGTTGTTAGGCATCAAGGCACC 8951
Qy 919 gaagtctctacgctgttggcacagctcgcttctggataaaagccagcgcgccattgatt 978
Db 8952 GAGATGAGTGCAGATATCGCCAAAGCCCTTCAACCTTGACGTGCAGCTGGCGCTTTGTC 9011
Qy 979 gccaaaatccttcccgagcccccgagaaacgctgcgcgcctgcagggcgagacatgc 1038
Db 9012 AGCGAAGTGTTCGCGAGGTTCTGGCTCGGCAAAAGCGCGCTCAAAAGCGCGCATATTAT 9071
Qy 1039 ctacgctcgacgcgagaaataacgttcttccgcgaccttcccgctcatgctcgcgcc 1098
Db 9072 ACCAGCTCAACGCAAAACCGCTGAATAGCTTTGCTGAGTTGCGGCTCTCGTATCGGACC 9131
Qy 1099 attacgcccggaaaaagatcagctcctgcgctgctatggcgaaggggaagaaatcacatc 1158
Db 9132 ACCGAGCGCGGCACGAAAGTGAAGCTTGCGCTGCTGCTAACGGCAAAACCACTGGAAGTA 9191
Qy 1159 aaagccaaagctgggcaacgcgcgc 1182
Db 9192 GAAGTGACGCTCGATACAGCACC 9215
```

RESULT 10
ECOUW67_1
WPCOMMENT

Sequence split into 4 fragments LOCUS ECOUW67 Accession U18997

Fragment Name	Begin	End
ECOUW67_0	1	110000
ECOUW67_1	100001	210000
ECOUW67_2	200001	310000
ECOUW67_3	300001	372438

Continuation (2 of 4) of ECOUW67 from base 100001 (U18997 Escherichia coli K-12 chrono

```
Query Match 14.1%; Score 196; DB 2; Length 110000;
Best Local Similarity 53.6%; Pred. No. 1.6e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;
```

```
Qy 382 ggcttcacacgcaaaacggctacatctgaccaataaccacgcgtgtcgcggtatg 441
- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61770 gtcacatcaacgccagtaaggctatgtctgaccaacacacccatcgattatcaggca 61829
```

QY 442 ggcagatcaaatcgtctgctcaacgacagcggaataatataccgcgaataatccgcaataatccggttcg 501
 Db 61830 CAGAAATAGTATTCAGCTCAATGATGGCGCGAGTTGTAGTCAAAACTGATTGGTAGC 61889
 QY 502 gatgtccaatccgatgctcgccctctgaaatcgacgcaacggaagagctaccgctgctc 561
 Db 61890 GATGACGAGCGATATCGCCCTGTACAAATTCAAAACCGCAGCAATTAACGCAATC 61949
 QY 562 aaaaatcggaatcccccaaaatttgaacccggcggaatgggtcgctgcatccggtcgccgccc 621
 Db 61950 GCTATTGCCACTCCGATAAATTCGCGTGGTATTTGGCTAGCGGTGGTAACCCA 62009
 QY 622 ttccggtttgaacacagcgtagcccgccgcatcgctccgcgaagcagaagcctgccc 681
 Db 62010 TTTGGCTTGGGCAACCGCACCTCTGGCATTTTCCGATTATTTCCGCGAGCGGGTTG 62069
 QY 682 aa---cgaaagctacacaccccttcacacacccgacgctgcatcaatccgggcaattcc 738
 Db 62070 AATCTTGAAGGTCTGGAAACTTTATCCAGACAGATGCTTCCATTACCGCGGTAACTCC 62129
 QY 739 ggcggcccgctgtcaacttaaaaggacagctgctggcatcaattcgcaaatatacagc 798
 Db 62130 GCGGTGCACTATTAAACCTTAACGGTGAGTTAATGGCATCAACACTGCAATCCTTGGC 62189
 QY 799 cgcagcggcgattcatgggcatctcctttgcatcccgattgacgttgccatgaatgctc 858
 Db 62190 CTTGGCGGGGAGCGTCGGGATTGGATTGGCCATCCCACTAATATATGGCGCAACACTG 62249
 QY 859 gccgaacagctgaaaaacacacgcgcgaagtcacacgcgcgaacactggcgctgattatcag 918
 Db 62250 GCCCAGCAGCTTATCGACTTTGGTGAATCAACGCGGTTGTAGCATCAAAAGGCACC 62309
 QY 919 gaagtatccatcaggtttggcacagctgctgggttgatataaagccagcggcgcatgatt 978
 Db 62310 GAGATGAGTCCGATATCGCAAGCCTTCAACCTTGACGTGCAGCTGGCGCGTTTGTGTC 62369
 QY 979 gccaaatcttccggcagccgcgcgaacagctgcccggcctgagcggcgagcagatgctc 1038
 Db 62370 AGCGAAGTGTTCAGGTTCTGGCTCGGCAAAACGCGGCTCAAAAGCGCGGATATTATT 62429
 QY 1039 ctgagctcgacggcgagaaaacgcttctccggcgacattccctgctgctggtcgcgccc 1098
 Db 62430 ACCAGCTCAACGGCAACCGCTGAATAGCTTGTCTGAGTTGCGCTCTGCTATCGCGACC 62489
 QY 1099 attacccggggaagaagtcagcctcgctgctggtgagcgaagcgaagaatcacatc 1158
 Db 62490 ACCGAGCGGGCACGAAAGTGAAGCTTGGCTGTGGGTAAACGGCAAAACCACTGGAAGTA 62549
 QY 1159 aaagccaagctggcgaacgcggc 1182
 Db 62550 GAAGTGACGCTCGATACCGACACC 62573

RESULT 11
 PAU29172
 LOCUS PAU29172 1191 bp DNA BCT 06-MAR-1996
 DEFINITION Pseudomonas aeruginosa HtrA-like serine protease Algw gene,
 complete cds.
 ACCESSION U29172
 VERSION U29172.1 GI:1184674
 KEYWORDS
 SOURCE Pseudomonas aeruginosa.
 ORGANISM Pseudomonas aeruginosa.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 Pseudomonas.
 REFERENCE 1 (bases 1 to 1191)
 AUTHORS Boucher,J.C., Martinez-Salazar,J., Schurr,M.J., Mudd,M.H., Yu,H.
 and Deretic,V
 TITLE Two distinct loci affecting conversion to mucoidy in Pseudomonas
 aeruginosa in cystic fibrosis encode homologs of the serine
 protease HtrA
 J. Bacteriol. 178 (2), 511-523 (1996)

96134987

MEDLINE 2 (bases 1 to 1191)

REFERENCE Boucher,J.C.

AUTHORS Direct Submission

TITLE Submitted (14-JUN-1995)

JOURNAL John C. Boucher, Microbiology, UTHSCSA,

7703 Floyd Curl Dr., San Antonio, TX 78284, USA

FEATURES Location/Qualifiers

source 1..1191

/organism="Pseudomonas aeruginosa"

/strain="PAO1"

/db_xref="taxon:287"

/map="69 min"

14..1183

/note="HtrA-like serine protease"

/codon_start=1

/transl_table=11

/product="Algw"

/protein_id="AAC43674.1"

/db_xref="GI:1184675"

/translation="MPRAFLFGLPVLGVLLALLIIQHNPVLGVPQEVHVEQAPL

LSRQEGPVSYANAVSRAAPAVANLYTTKMKRPSHPLFDDPMFRFFGDNLPQOKRM

ESSLSAVIMSAGYLLTNHVTAGADIIVALRDGRETIAQLVGDPTDIAVLKID

LKNPAMTIGRSDGIRTDVCLAIKNPFGVGTDRDGHORHRTOPARDVTYEDFIOT

DAAINPGNSGALVDAAGNLIGINTAIFSKSGSOGIRSPSRPSTLEVMQSIIEHQ

VIRGWLGVKALTPELAESLIGETAGIVVAGRLSRGPAARGLLPGDVLITDKQE

ASDGRSMNQVARTPGQKISIVLRNGQKVNLTAEVGLRPPPPAPQOKDQGE"

BASE COUNT 214 a 428 c 376 g 173 t

ORIGIN

Query Match

Best Local Similarity 13.4%; Score 186.6; DB 2; Length 1191;

Matches 462; Conservative 54.4%; Pred. No. 1.4e-28;

Mismatches 379; Indels 8; Gaps 4;

QY 367 ttgaacttcggttcgggttcctcatcagcaaaacgcgctacatcctcgaccaatccacc 426

Db 323 TCGAGCCTCGGCTCGGCGGTGATCATGAGCGCGGAGGCTACTGCTGACCAACACACCAC 382

QY 427 gtctgttcggttatgggcagtatcaaatcctgctcaacgacaagcgcgaatataccgc 486

Db 383 GTGACCGCTGCGCGCACAGATCATCTGTGGCTTTGCGCGACGGCGCGCAACCATCGCC 442

QY 487 aaatcctcgttcggatgtccaatcccgatgtcgccctcttgaataatcgacgaaggaa 546

Db 443 CAGTTGGTCGCGACGACCGCAACCGACCTGGCGCTGCTGAAGATCGACCTTA---AG 499

QY 547 gagctaccctcgctcaaaatcggaatcccaaaatttgaaacccggcggaatggcgct 606

Db 500 AACCTGCGCGGATGACCTTCGGCGCTCCGACGCGATTCGACCGCGGACGCTCTGCCCTC 559

QY 607 gccatcgcgcccttcggcttttgacaaca--gctgacccgcggcagctgctcgcccaa 665

Db 560 GCCATCGCAACCCGTTCCGGGTGCGCGACACCCGTGACCATGGGCTCATCAGACCGCGCGATC 619

QY 666 aggcagaagctgcccac--gaaagctacacaccccttcacccaacacgagctgcatc 723

Db 620 CGGACGCAACAGCTTCGGCTGACACCTACGAAAGATTCATCCAGACCGACGCGCGATC 679

QY 724 aatccgggcaattccggcgcccgctgtcaacttaaaaggacaggtcgctcgacatcaat 783

Db 680 AACCCGGCAACTCCGGCGCGCGCTGGTGGACGCTGCGCGCAACCTGATCGGCATCAAC 739

QY 784 tcgcaaatatacagccgagcgggattcatgtgggcatctcttggccatcccgattgac 843

Db 740 ACGGCGATCTTCTCCAAAGTCGGCGCGCTCCAGGGTATCCGTTCCGCTATCCGACCAAGC 799

QY 844 gttccatgaatgctgcgcaaacagctgaaacacacgcgcaaaagtcacacgagacaatcg 903

Db 800 TGGACCCCTGGAGGTGATCATGTCGATCATCGAGACGACGCCAGGTCCTCGGCTCGGCTC 859

QY 904 ggcgtgattatcaggaagtatcctacggtttggcacagctgctcggttctggtgataagcc 963

Db 860 GCGTCGAGGTCGAAGCGCTGACCCCGGAACTGGCGGAGTGGCGCTCGGCGCAACACC 919

- VERSION D26562.1 GI:473770
KEYWORDS 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase; GMP reductase; PII uridylyl-transferase; adenosylmethionine decarboxylase; proenzyme; ampd signalling protein; ampe signalling protein; aromatic amino acid transport protein; deoxyguanosine triphosphate triphosphohydrolase; dihydroliipoamide acetyltransferase; dihydroliipoamide dehydrogenase; dosage-dependent dnaK suppressor protein; elongation factor EF-Ts; ferrichrome-iron receptor; ferrichrome-iron transport protein fhb; ferrichrome-iron transport protein fhuc; ferrichrome-iron transport protein fhud; glucose dehydrogenase; glutamate-1-semialdehyde aminotransferase; heat shock protein hscA; methionine aminopeptidase; mukB suppressor protein snbA; mutator mutR; nicotinate-nucleotide pyrophosphorylase (carboxylating); penicillin-binding protein 1B; plasmid copy number control protein pcnB; pyruvate dehydrogenase; regulatory protein sfsl; ribosomal protein S2; ribosome releasing factor; spermidine synthase; tetrahydridipicolinate N-succinyltransferase; Escherichia coli (sub-strain W3110, strain K-12) (library: Kohara's mini set library) DNA, clone #111, 112, 113, 114, 115, 116 and 119. Escherichia coli Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
- REFERENCE 1 (bases 1 to 82727)
AUTHORS Fujita, N.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-1994) to the DDBJ/EMBL/GenBank databases.
Nobuyuki Fujita, National Institute of Genetics, Department of Molecular Genetics; 1111 Yata, Mishima, Shizuoka 411, Japan (E-mail: nfuji@nig.ac.jp, Tel: +81-559-75-0771 (ex.684), Fax: +81-559-71-3651)
- REFERENCE 2 (sites)
AUTHORS An, G., Bendtak, D.S., Mamelak, L.A. and Friesen, J.D.
TITLE Organization and nucleotide sequence of a new ribosomal operon in Escherichia coli containing the genes for ribosomal protein S2 and elongation factor Ts
JOURNAL Nucleic Acids Res. 9 (16), 4163-4172 (1981)
MEDLINE 82059454
REFERENCE 3 (sites)
AUTHORS Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J.R.
TITLE The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the pyruvate dehydrogenase component
Eur. J. Biochem. 133 (1), 155-162 (1983)
JOURNAL 83209630
MEDLINE 4 (sites)
AUTHORS Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J.R.
TITLE The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the dihydroliipoamide acetyltransferase component
Eur. J. Biochem. 133 (3), 481-489 (1983)
JOURNAL 83234434
REFERENCE 5 (sites)
AUTHORS Stephens, P.E., Lewis, H.M., Darlison, M.G. and Guest, J.R.
TITLE Nucleotide sequence of the liipoamide dehydrogenase gene of Escherichia coli K12
Eur. J. Biochem. 135 (3), 519-527 (1983)
JOURNAL 84004369
MEDLINE 6 (sites)
AUTHORS Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J.C.
TITLE Regulation of expression and nucleotide sequence of the Escherichia coli gapD gene
J. Biol. Chem. 259 (23), 14824-14828 (1984)
JOURNAL 85054973
MEDLINE 7 (sites)
AUTHORS Broome-Smith, J.K., Edelman, A., Yousif, S. and Spratt, B.G.
TITLE The nucleotide sequences of the ponA and ponB genes encoding penicillin-binding protein 1A and 1B of Escherichia coli K12
Eur. J. Biochem. 147 (2), 437-446 (1985)
JOURNAL 85127060
MEDLINE 8 (sites)
AUTHORS Coulton, J.W., Mason, P., Cameron, D.R., Carmel, G., Jean, R. and Rodé, H.N.
TITLE Protein fusions of beta-galactosidase to the ferrichrome-iron receptor of Escherichia coli K-12
- JOURNAL 165 (1), 181-192 (1986)
MEDLINE 86083668
REFERENCE 9 (sites)
AUTHORS Breton, R., Sanfacon, H., Papayannopoulos, I., Biemann, K. and Lapointe, J.
TITLE Glutamyl-tRNA synthetase of Escherichia coli. Isolation and primary structure of the gltX gene and homology with other aminoacyl-tRNA synthetases
J. Biol. Chem. 261 (23), 10610-10617 (1986)
JOURNAL 86278132
MEDLINE 10 (sites)
AUTHORS Koster, W. and Braun, V.
TITLE Iron hydroxamate transport of Escherichia coli: nucleotide sequence of the fhbB gene and identification of the protein
Mol. Gen. Genet. 204 (3), 435-442 (1986)
JOURNAL 87014116
MEDLINE 11 (sites)
AUTHORS Chye, M.L. and Pittard, J.
TITLE Transcription control of the aroP gene in Escherichia coli K-12: analysis of operator mutants
J. Bacteriol. 169 (1), 386-393 (1987)
JOURNAL 87083395
MEDLINE 12 (sites)
AUTHORS Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A. and Chang, S.
TITLE Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure
J. Bacteriol. 169 (2), 751-757 (1987)
JOURNAL 87109068
MEDLINE 13 (sites)
AUTHORS Coulton, J.W., Mason, P. and Allatt, D.D.
TITLE fhuc and fhud genes for iron (III)-ferrichrome transport into Escherichia coli K-12
J. Bacteriol. 169 (8), 3844-3849 (1987)
JOURNAL 87279948
MEDLINE 14 (sites)
AUTHORS Tabor, C.W. and Tabor, H.
TITLE The speE operon of Escherichia coli. Formation and processing of a proenzyme form of S-adenosylmethionine decarboxylase
J. Biol. Chem. 262 (33), 16037-16040 (1987)
JOURNAL 88058963
MEDLINE 15 (sites)
AUTHORS Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wächter, E.
TITLE Complementary DNA and derived amino acid sequence of the precursor of one of the three protein components of the inter-alpha-trypsin inhibitor complex
FEBS Lett. 229 (1), 63-67 (1988)
JOURNAL 88152237
MEDLINE 16 (sites)
AUTHORS Andrews, S.C. and Guest, J.R.
TITLE Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12
Biochem. J. 255 (1), 35-43 (1988)
JOURNAL 89061679
MEDLINE 17 (sites)
AUTHORS Mellano, M.A. and Cooksey, D.A.
TITLE Nucleotide sequence and organization of copper resistance genes from Pseudomonas syringae pv. tomato
J. Bacteriol. 170 (6), 2879-2883 (1988)
JOURNAL 88227880
MEDLINE 18 (sites)
AUTHORS Liu, J.D. and Parkinson, J.S.
TITLE Genetics and sequence analysis of the pcnB locus, an Escherichia coli gene involved in plasmid copy number control
J. Bacteriol. 171 (3), 1254-1261 (1989)
JOURNAL 89155419
MEDLINE 19 (sites)
AUTHORS Sung, Y.C. and Fuchs, J.A.
TITLE Characterization of the cyn operon in Escherichia coli K12
J. Biol. Chem. 263 (29), 14769-14775 (1988)
JOURNAL 89008347
MEDLINE 20 (sites)


```

/db_xref="taxon:562"
/clone_lib="Sequencing library prepared from pulsed field
gel purification of genomic AvrII fragments of E. coli
strain MGL655 provided by Fred Blattner"
<1..802
/gene="fhuA"
/codon_start=2
/transl_table=11
/product="ferrichrome-iron receptor precursor"
/protein_id="AAB08580.1"
/db_xref="GI:1552728"
/translation="LGGRYDADQESLNRVAGTDDKRDQKQFTHRGGVNYLFDNGVTP
YFSYSESPSQKGDNIAPSKQYQEVGKRVYVTPEDRPIVVTGNTKNNLM
ADPEGSEFVSGEIRAGVEIEAKAALSASVNVVGYSTYDDATDTTDTTKGNTPAQ
VPKHMASLWADYTFDGLSLGLTGRTGRTGSSGYDPANSPKFGSYTVVDALRYDL
ARVGMAGSNVALHVNLFDEVVASCFNTYGCFWGAERQVATATFRF"
1..802
/gene="fhuA"
853..1650
/gene="fhuC"
853..1650
/gene="fhuC"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuC"
/protein_id="AAB08581.1"
/db_xref="GI:1552729"
/translation="MQEYTNHSDTTFALRNISFRVPGRTLLHPLSLTPAGKVTGLIG
HNGSKSLIKMLGRHOPPSGEILLDAQLESWSKAPARKVAYLPOLPAEGMTV
RELVAIGRPWHGALGRFGAADREKVEEASISVLGKLPLAHLRYDLSLSEGERQRIAM
LVAQSRCLLDEPTSLDIAHQVDVLSLVHLSQERGLTVAVLHDHINMAYRCDYL
VALRGEMIAQGTAPAEIMRGETLEMIYGIPMGPLPHPAGAEPVSFVY"
1650..2540
/gene="fhuD"
1650..2540
/gene="fhuD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuD"
/protein_id="AAB08582.1"
/db_xref="GI:1552730"
/translation="MSGPLISRRLLTAMALSPILLQWNTAHAAIDPNRIVALEWL
PVELLALGIVPYGVADTINRLWSEPPDPDSVIDVGLRTEPNLELLEMPSEMYW
SAGYGPSEMARIAAPGRGNFSDQKPLAMAKSLTEMADLLNLQSAATHLAQYED
FTRMKPRFKRGARPLLTLLDPRHMLVFGPNSLFQELDEYGPINAQOGETNFG
STAVSDRLAAYKQDVCLCFDHDNSKMDMALMATPLWQAMPEVRAGRFQVRPAWVFY
ATLSAMHFVRVLDAIGGKA"
2537..4519
/gene="fhuB"
2537..4519
/gene="fhuB"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuB"
/protein_id="AAB08583.1"
/db_xref="GI:1552731"
/translation="MSKRIALFPALLLALLIVATALTWMNFSQALPRSQWAQAWSP
DIDVQIMTFHYSLLPRLAISLVAGAGLVGLVFOQLRNPPLAEPTTLGVATGAQLG
ITVTTLWAIPEGMASOFAAQAGACVGLIVFGVWGRKLSPTVLLIAGLVSLYCGAI
QALLVIFHDQOQSLMSTGTLTQTDWGVVERLWQOLLGGVMTLLRLPLTLMGLD
DGVNRLGALSILUARLAULSILVIALVNVAGIIGIFGLFAPLAKMLGARLLPR
LMLASIGALILWLSQIITLWTRVWMEVSTGVTALIGAPLLLLPLRLRSLSAPDM
KNDRVAARQHVHAFALAGVILLMAVVYALSGRDAGHGTWASGALLDLMPWEP
RIMAAFGVMLAVAGCIIQRLGNPMASPEVLGISSGAFGVLLMLFLVPGNGWL
LPAGSLGAATLLIIMAAARGGCFSPHMLLAGMALSTAFMLLMLOAGSDPRMAQV
LWISGSTYNATDAQVRGIVNVILLATPCLRRWLTILPLGDDIARAVGMAITPR
IALLLLAACLTATMTTGPLSEFVGLMAPHIARMGFRTRMPHIVISALVGGLLFLFA
DMCGRMVLPFPQIPAGLSTFGIAPFYIYLLRQSR"
complement(4677..5957)
/gene="gsa"

```

```

CDS
complement(4677..5957)
/gene="gsa"
/codon_start=1
/transl_table=11
/product="glutamine-1-semialdehyde aminotransferase"
/protein_id="AAB08584.1"
/db_xref="GI:1552732"
/translation="MSKSENLYSAARELIPGVNVPYRAFTGVGTPLFIEKADGAYL
YVDGKAYIDVSGMPVGLGNHPALNRNAVIEAERGLSFGATEMVEVKAQLYTEL
QPTNDVMVNSGTEATMSAIRLARGFTGRDKIIFKFCGYGHGADCLLVKAGSGALTLL
GPNMSPGVADFAKYTLTCTYNDLASVRAAFQYQEIACIIVFVAGNMNCVPPLE
FLGRDALCDEFGALLIDEVMTGRVALAQAQDYGVVVDLTCIGIKGGMPVLAEG
GGRVDMDALAPTGVVQAQTLSCGNLAAAGACLNVAOPGVHETIDELTTRLAEG
LLEAAERAGILPVNVHVGCMFGIIFTDAESTYCTQDVNACDVERKRFKRFHMLDSEVY
LAPSAFAGFMSVAHSMEDINNTIDAARVFAKL"
6182..7603
/notes="similar to voltage-gated chloride channel protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08585.1"
/db_xref="GI:1552733"
/translation="MKTDTPSLETPOAARLRRLRIQLERDKTPLAILFMAAVVGT
LVGLAAVAFDKGVALEDORPVRWVLPVKFFGGLTGLGGMVLRGEPVQVGGNIGR
EAGSGIPEIGALEDORPVRWVLPVKFFGGLTGLGGMVLRGEPVQVGGNIGR
MYLDIFRLKGEARHTLLATGAAGLAAAFNAPLAGILFIIEEMRQPRYTLISIKAV
FIGVMSTIMIRIFNHEVALIDVKGKLSAPLNTLWLYLILGLIFIGFPIENKWLGM
QDLHRVHGNNITKWLMMGAIGGCLGLGVAPATSGGGFNLIPIATAGNFSMGLV
FIFVARVTTILCFSSGAGGIFAPMLALGTVLGFAGMVAVELPFQVHLEAGTFAIA
GMGLAAASIRAPLTGIILVLEMTDNYQLILPMIITGLGATLLAQFTGCKPLYSAILA
RTLAKQAEQQLARSKAASASENT"
7685..8029
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08586.1"
/db_xref="GI:1552734"
/translation="MSDDVALPLETDAANKVKSIIADENPNLKLRYITGGGCSG
FOYGFETDQVNEGDMTEKQGVGLVDPMSLOQLVGGSDVTEGLEGSRFVITNPN
KSTCGGSSFSI"
complement(8076..8699)
/notes="similar to Vibrio cholerae RecA protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08587.1"
/db_xref="GI:1552735"
/translation="MLVYWDIVGTAVFAISGVLLAGKLRMDPFGVYLVGVTVAGGG
TIRDMALDHGPVFWKDPDLDLVAVMTSMLTIVLVRQPRRLPKMMLPVLDAVLAVFV
GIGNKAFNAEAGPLIAYCMGVITGVGGIIRDVLAREIPMLIRTEIYATACIIGIV
HATAYTFSVPLETASMMGMVYLLRLAARHKLKLPFDLDENGR"
complement(8737..9537)
/gene="yadT"
complement(8737..9537)
/gene="yadT"
/notes="hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08588.1"
/db_xref="GI:1552736"
/translation="MAKSLFRALVSLFPLNLNAPRVITLSPANTELAFAAGITTP
YGVSSYSDPPROAKIEOVSTWQGNLERIVALKPDLVIWARGGNARQVDQASLGI
KVMVDNATSEIQANALRQLAPSPQPDKAQNAQSLDQYLAQLKQYADKPKRVFL
QWGNPFPTSGESIQNVLEWCGGENIFRSDQPPWPQVRSQVRSQVRSQVRSQVRSQV
DQIPKTKWGEQLKIPVLPITSDMFERASPRILAAQQLCNALSQVD"
complement(9530..10228)
/notes="similar to purine nucleoside phosphorylase (deod)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08589.1"

```

```
/db_xref="GI:1552737"
/translation="MKIIGIAMEEEVTLIRDKIENRQTISLGCEIYTQQLNGTEVA
LKSGIGKVAALAGLALLLHCHRPDVIINTGSAGGLAPTLUKVGDIVVSDRHADY
TAFYGLPGCPAGFKADKLIAAEACIAELNNAVRGLIVSGDAFINSGYGLAK
IRHNFPAIAVEATEATAIHVCHNFNVFVVVRAISDVADQSHLSDFELAVAAKQS
SLMVESLVOKLAHG"
10312..11829
/gene="dgt"
CDS
10312..11829
/gene="dgt"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="deoxyuanosine triphosphate"
/protein_id="AA808590.1"
/db_xref="GI:1552738"
/translation="MAQIDFRKKNWHRYSPOGVKTEHILRIFESDRGRINSFA
IRLQOKTOVPELERNAAVTRLTHSMEVOQVGRYIAKETLSRLKLLKLEAYGLDEL
TGFESIVEMSLMHLIDIGNPPFGHGEAALNDFRQLHPEDAESQPLTDRGSVAAL
RGTGEPLNELRRKIRQDLCHFEGNAQGRILVHTLRMNLTWQAQGGILKLYTRPAWM
RGTPTPHLLMKRPGYLYSEAYIARLKEKLNALYSREPLTWIMEAADDISYCVAD
LEDAVEKRITFVQLYHHLHEAWGOHEKGSFLSLVVENAWEKSRNSLSRSTEDQFM
Query Match 12.8%; Score 178; DB 1; Length 123171;
Best Local Similarity 52.1%; Pred. No. 7.1e-27;
Matches 422; Conservative 0; Mismatches 385; Indels 3; Gaps 1;
QY 382 ggttcattcatcagcaaaacggctacatcctgaccataaccacgcgtgttgcgggtatg 441
Db 12307 GTCATATTGATCGGATAAAGGCTATGCTGTCACCAACACACACGTTGTTGATAACGCG 12366
QY 442 ggcagtatcaagtctctgctcaacgacaaagcgcgaaatataccgcgaactcatcggttcg 501
Db 12367 ACGGTGATTAAAGTTCAACTGAGCGATGGCCGTAAGTTCGACGCGAAGATGGTTGGCAAA 12426
QY 502 gatgtccaatccgatgctgcctcttgaatacgacgcaacgaagagctaccgcgtgctc 561
Db 12427 GATCCGGCTCTGATTCGGCTGATCAATCAATCAGAACCGGAAACCTGACCGCAATT 12486
QY 562 aaaaatcggaatccccaaaaatttgaaacccgggggaattgggtcgctgcattcgccgcgcc 621
Db 12487 AAGATGGCGGATCTGATGCACCTGCGGTGGTGATTACACCGTAGCGATTGGTAAACCCG 12546
QY 622 ttccgctttgacaacagctgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 678
Db 12547 TTTGGTCTGGCGAGACGGTAACTTCGGGATTTCTCTGCGCTGGGCGTAGCGGCGCTG 12606
QY 679 cccaagaaagctacacaccccttcacccaaacgcagcttgccatcaatccgggcaattcc 738
Db 12607 AATGCCGAAACTACGAAACTTCATCCAGACCGATGCGAGGATCAACCGTGGTAACTCC 12666
QY 739 ggcggcccgctgttcaacttaaaagacaggtgtcgtgcgatcaattcgcaaatatacagc 798
Db 12667 GGTGGTGGCTGTTTAACTTGAACGGGAACTGATCGGTATCAACACCGCATCTCGCA 12726
QY 799 cgcagcgcgcgattcattgagcattctcttgcacatcccgattgacgttgccatgaatgc 858
Db 12727 CCGAGCGCGCGCAATCTCGGTATCGGTTTGTCTATCCGAGTACATGGTGAACAACTG 12786
QY 859 gccgaacagctgaaaaacacccgcgaagtcacacgcgcggaacattggcggtgattattcag 918
Db 12787 ACCTCGCAGATGTTGGAATACGGCCAGGTGAACACGGGTGAGCTGGGTATTATGGGACT 12846
QY 919 gaagtatctcaggtttggaacagtcgttcggtctgataaaagccagcgcgcatattgatt 978
Db 12847 GAGCTGAACCTCCGAACCTGGCGAAGCGCATGAAAGTTGACGCCCGCGCTGCTTCGTA 12906
QY 979 gccaaaatcttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1038
Db 12907 AGCAGGTTCTGCCTAATTCCTCGCTGCAAAACCGGGCATTAAGCGGGGTGATGTGATC 12966
QY 1039 ctacgacctcagcggcgagaaatacgttcttccggcgaccttcccgctcatggtcggcgcc 1098
```

```
Db 12967 ACCTCACTGAACGGTAAGCCGATCAGCAGCTTTGCCGCACTGGCTGCTCAGGTGGGTACT 13026
QY 1099 attacgcgggaaaaaagtagtcagctcgcgctatgctgcgcgaagcgcaagaaatcacaatc 1158
Db 13027 ATGCGGTAGGCAGCAAACTGACCTGGCTTACTGCGGACGGTAAGCAGGTTAACGTG 13086
QY 1159 aaagccaaagctgggcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1188
Db 13087 AACCTGGAACCTGCAGCAGCAGCAGCAGCAAT 13116
RESULT 15
ECHTRA 1855 bp DNA BCT 05-JUL-1999
DEFINITION Escherichia coli htra gene for 51kd protein.
ACCESSION X12457
VERSION X12457.1 GI:41760
KEYWORDS htra gene.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1855)
AUTHORS Lipinska,B. and Sharma,S.
TITLE Sequence analysis and regulation of the htra gene of Escherichia coli: A sigma-32-independent mechanism of heat-inducible transcription
JOURNAL Nucleic Acids Res. 16, 10053-10067 (1988)
MEDLINE 89057448
REFERENCE 2 (bases 1 to 1855)
AUTHORS Lipinska,B.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1988) Lipinska B., CUMB Dept., University of Utah, Salt Lake City, Utah 84132, USA
FEATURES
source 1..1855
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
/clone="pBL52"
/map="4 min.between tonA and dapD gene"
111..159
/note="promoter region homologous to the E. coli P3 promoter of rpoH gene"
misc_feature 159
misc_feature /note="transcription initiation site"
199..1674
CDS /note="htra protein (AA 1-491)"
/codon_start=1
/transl_table=11
/protein_id="CAA30997.1"
/db_xref="GI:41761"
/db_xref="SWISS-PROT:P09376"
/translation="MKKTTLALSRLSLALSPLSATAAFTSSATTAQOPLSLAPM
LEKVPVSVINVEGSTVTNTPRPNFOFEGDDSPFCOBGSPFOSSPFCOGGOGGN
GGGQQQFMALGSGVVIDADKGVYVTVNNHVVNDATVIVQSDGRKFAKVMYKQDPRS
DIALIQNPNKLTAKMADSDALRVGDTYVIGNPFGLGETVTSGIVSALGRSLNA
ENYENFTDAAINRNGSGALVNLNGELIGINTAILADPGDNIGTIGFAPSNMYKNL
TSQVVEYGOVKRGELGIMTELSLAKAMKVDQAFVSOVLNPSAAAKAGIKAGD
VITSLNGKPISSFAALRAQVMPVGSKLITLGLLDGKVNVNLEQQOSSONQVDSST
IFNGIEAEMSNKGDQGVYNNVYKGTTPAAQIGLKGDVILIGANQAVKNTAELRKV
LDSFVSLALNIQRDLHPVNAVLSNPLFKTGRGSPYL"
```

```
BASE COUNT 480 a 467 c 491 g 417 t
ORIGIN
Query Match 12.6%; Score 176.4; DB 1; Length 1855;
Best Local Similarity 52.0%; Pred. No. 1.6e-26;
Matches 421; Conservative 0; Mismatches 386; Indels 3; Gaps 1;
QY 382 ggttcattcatcagcaaaacggctacatcctgaccataaccacgcgtgttgcgggtatg 441
Db 547 GTCATCATTTGATCGCGATAAAGGCTATGTCGTCAACCAACACACGCTTCTTGATAACGCG 606
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:51:56 ; Search time 120.64 Seconds
(without alignments)
2893.053 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 ggttcaaaaataccaata.....tcagggaagtcctccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	176.4	12.6	1854	1 V29528	Escherichia coli s
2	175.6	12.6	1980	1 Q14116	S. typhimurium htr
3	154.2	11.1	1791	1 T04403	Heat shock protein
4	139.4	10.0	2894	1 T10423	H. influenzae S833
5	128.8	9.2	1529	1 X14164	H. pylori GHFO 536
6	124	8.9	1540	1 V30845	Nucleotide sequenc
7	120.8	8.7	1484	1 V30921	Nucleotide sequenc
8	110.8	7.9	558	1 V31239	E. coli J96 pathog
9	105	7.5	2036	1 V29540	Homo sapiens D8725
10	105	7.5	2075	1 T75444	CDNA encoding oste
11	105	7.5	2205	1 V04680	Human presenilin 1
12	85	6.1	607	1 X29130	Polynucleotide RNP
13	71	5.1	284	1 V90593	Nucleotide sequenc
14	68	4.9	1553	1 V43618	Human secreted pro
15	62.8	4.5	273	1 V90690	Nucleotide sequenc
16	60.6	4.3	15747	1 X13018	Enterococcus faeca
17	60.6	4.3	21338	1 X52153	Streptococcus pneu
18	60.4	4.3	9848	1 X20503	Polynucleotide seq
19	57.6	4.1	1835	1 V29524	Homo sapiens PSP1
20	57.6	4.1	2040	1 V08860	Human cancer-relat
21	57.6	4.1	2187	1 V29535	Homo sapiens PSP1
22	57.6	4.1	2187	1 V29536	Homo sapiens PSP1
23	57.6	4.1	2187	1 V29539	Homo sapiens PSP1
24	57.4	4.1	1395	1 X34246	Mycobacterium spec
25	57.4	4.1	1771	1 T91473	Mycobacterium tube
26	57.4	4.1	1771	1 T91410	Mycobacterium tube
27	57.4	4.1	1771	1 V44351	Mycobacterium tube
28	57.4	4.1	1771	1 V64459	M. tuberculosis im
29	57.4	4.1	1812	1 X34247	Mycobacterium spec
30	57.4	4.1	2162	1 V37380	Streptococcus pneu
31	55	3.9	1506	1 X20639	Polynucleotide seq
32	51.6	3.7	1503	1 V29523	Homo sapiens PSP1
33	51.6	3.7	1787	1 V29522	Homo sapiens PSP1
34	51.6	3.7	1987	1 V74833	Staphylococcus aur

ALIGNMENTS

RESULT 1

V29528
ID V29528 standard; cDNA; 1854 BP.
AC V29528;
DT 13-OCT-1998 (first entry)
DE Escherichia coli serine protease htrA.
KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease; htrA;
KW serine protease; neurodegeneration; predisposition; diagnosis; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
CDS 199..1674
FT /*tag= a
FT /product= htrA serine protease
PE EP-828003-A2.
PD 11-MAR-1998.
PF 26-AUG-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,
PI Livi GP, Southan CD;
DR WPI; 98-161101/15.
DR P-PSDB; W56771.
PT Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
PS Example 2; Page 27-28; 65pp; English.
CC The sequence is that encoding the serine protease htrA which was used
CC in the isolation of PSP1.
SQ Sequence 1854 BP; 478 A; 470 C; 489 G; 417 T;

Query Match 12.6%; Score 176.4; DB 1; Length 1854;
Best Local Similarity 52.0%; Pred. No. 2e-38;
Matches 421; Conservative 0; Mismatches 386; Indels 3; Gaps 1;
QY 382 ggcttcacatcagcaaaacggctacatccctgacacataccacagctcgttgcggtatg 441
Db 547 GTCATCATTTGATGCGGATTAAGGCTATGTCGTCACCAACACCGTTTGTGATAACGCG 606
QY 442 ggcagatcaaaagtctctgctaacgacagcgaataataccgcaaacatcgcgttcg 501
Db 607 ACGGTCATTTAAAGTTCAACTGACGATGCCCGTAAAGTTTCGACGCGAAGATGGTTGCAAA 666
QY 502 gatgtccaatccgatgcgcctcttgaataacgacgcaacggaagagctaccctgcgtc 561
Db 667 GATCCCGGCTGTGATATCCGCTGTATCCAAATCCAGAACCCGAAACCTGACCGCAATT 726
QY 562 aaaaatcgcaatcccaaaaattgaaacccggcggaatgggtgcgtgcctatcgcgcgccc 621
Db 727 AAGATGCGGATTTCTGATGCTACCTGCGCTGGGTGATTACACCGTAGGATTTGTAACCCG 786
QY 622 ttccggtttgacaaacagctgacccgcggcatgttccgcgaag---gcagaagcctg 678
Db 787 TTTGTGCTGGCGGAGACGGTAACCTCCGGGATTTGCTCTGCTGGGGCGGTAGCGGCTG 846

Homo sapiens PSP1-
Homo sapiens PSP1-
Continuation (8 of
Continuation (9 of
Polynucleotide seq
Mycobacterium spec
Mycobacterium spec
Mycobacterium tube
Mycobacterium tube
M. tuberculosis im

QY 679 ccaacgaagctacacaccccttcacaaacacagcgttgccatcatcccgccgaattcc 738
 Db 847 ATGCCGAAACACTACGAAACTTCATCCAGACCAGTCAGCGATCAACCGTGGTAATCC 906
 QY 739 ggcggcccgctgttcaacttaaaagacagctgctggcgcacatcgaattgcgaatacagc 798
 Db 907 GGTGGTGGCTGTGTTAACTTGAACGGGAACTGATCGGTATCAACACCGCGATCTCGCA 966
 QY 799 cgcagcggcggtatcattggcgcattctcttggcaccatcccgatcgatgacgttgcgaatgctc 858
 Db 967 CCGGACGGCGGCAACATCGGTATCGGTGTTGCTATCCGAGTAACTGGTGAACAACTG 1026
 QY 859 gccgaacagctgaaacacccgcgaagtcacagtcgaacgagcaacactggcggaattattcag 918
 Db 1027 ACCTCCAGATGTGGAATACGGCCAGGTGAACCGCGGTGAGCTGGGTATATCGGGACT 1086
 QY 919 gaagtacctcagctgtttggcacagtgcttgcgttgatgataaaagccagcgcgcattgatt 978
 Db 1087 GAGCTGAACCTCCGAACTGGGGAAGCGATGAAGTTGACGCCACGCGGTGCTTTCGTA 1146
 QY 979 gccaaatactctcccgccagcccgccgacagctgcggcgctgcaggcggcgacatcgtc 1038
 Db 1147 AGCCAGGTTCTGCTTAATCTCCGTCGCAAAAGCGGGCATTTAAAGCGGTGATGTGATC 1206
 QY 1039 ctacgctcgacggcgagaaatacgttcttcggcgacacttcccgatcagtggtggcgcc 1098
 Db 1207 ACTCTACTGAACGGTGAAGCGATCAGCAGCTTTGCCGCATCGGTGCTCAGGTGGGTACT 1266
 QY 1099 attacccggggaagaagtcagcctcggcgatggtgcgcaaaagcggaagaaacacacatc 1158
 Db 1267 ATCCCGGTAGGACGAACTGACCTCGGGTTACTGCGCGCAGGTAGCAGGTAAAGTG 1326
 QY 1159 aaagccaaagctgggcaacggccgagcagcat 1188
 Db 1327 AACCTGGAAGTCCGCGAGAGCAGCCAGAAT 1356

RESULT 2

Q14416
 ID Q14416 standard; DNA; 1980 BP.
 AC Q14416;
 DT 14-JAN-1992 (first entry)
 DE S. typhimurium htrA gene.
 KW TnpA mutagenesis; signal peptide; stress protein;
 KW heat shock protein; degp; ss.
 OS Salmonella typhimurium strain C5.
 FH Key Location/Qualifiers
 FT cds 395..1819
 FT FT /*tag= a
 FT /label= HtrA
 FT /note= "heat shock protein"
 FT 1833..1855
 FT FT /*tag= b
 FT /label= dyad_symmetry?
 FT /note= "not labelled in specification"
 PN W09115572-A.
 PD 17-OCT-1991.
 PF 23-MAR-1991; G00484.
 PR 30-MAR-1990; GB-007194.
 PA (WELL) WELLCOME FOUNDATION LTD.
 PI Dougan G, Charles IG, Hormaeche CE, Johnson KS, Chatfield SN;
 DR WPI: 91-325215/44.
 DR P-PSDB: R14625.
 PT Attenuated microorganism useful in live vaccines - attenuated by
 PT mutation in DNA sequence encoding e.g. a heat shock protein
 PS Disclosure; Fig 1; 28pp; English.
 CC TnpA mutagenesis was used in the mouse virulent S. typhimurium
 CC strain C5. Mutants were selected likely to harbour lesions in genes
 CC that have a signal peptide sequence, i.e. proteins likely to be
 CC targeted through a bacterial membrane. Isolation of the DNA
 CC flanking the tnpA insertion identified the gene (htrA) that has
 CC been insertionally activated. Comparison of the translated protein
 CC sequence showed that it shared 88% homology with the sequence of the

CC htrA product from E. coli.
 SQ Sequence 1980 BP; 482 A; 481 C; 551 G; 466 T;
 Query Match 12.6%; Score 175.6; DB 1; Length 1980;
 Best Local Similarity 52.6%; Pred. No. 3.3e-38;
 Matches 407; Conservative 0; Mismatches 364; Indels 3; Gaps 1;
 QY 400 aacggctacatccctgaccacatcaccacgctgttgcggtatgggagtgatcaaaagtcctg 459
 Db 764 AAGGGCTACGTCGTCACCAACACACACGCTGTGTAACGCCAGCGTGATTAAGGTACAG 823
 QY 460 ctcaacgacaaagcgcgaatacaccgcaaaactcaccgttcggttcggaatgccaatcccgatgc 519
 Db 824 CTTACGGATGGCGTAAATTCGATCTAAAGTGGTGGCAAGATCCCGGTTCTTGATATC 883
 QY 520 gccctcttgaataatcgacgcaacggaagactaccctcgctcaaaatcgccaatcccaaa 579
 Db 884 CGCTGATTCAAATTCAGAAATCCGGAAGAACCTGACGGGATTAAGCTGCGGACTCCGAC 943
 QY 580 aattgaaacggcggaatgggtcgctgccatcgcgcgcccttcggtttgacaaacgc 639
 Db 944 CGGCTGCGGTGGGGATTAACCTGCTATTGCTAAACCCGTTTGGTCTGGCGAAACG 1003
 QY 640 gtgaccgcgcatcgtgtccgccaag---gcagaagccttgcccaacgaagactacaca 696
 Db 1004 GTGACGTAGATATCGTTTCGGCGCTGGGGCTAGCGCCTGAAGCTAGAAAATACGAG 1063
 QY 697 cccttaccacaaacgacgttgccatcccgggcaattccggcgcccgctgttccaac 756
 Db 1064 AACTTTATTCAGACCGACGCGCGATTAACCGTGTGTAACCTCGCGCGCGCTGGTGAAC 1123
 QY 757 ttaaaagacaggtcgtcgccatcaattcgcaaatatatacagcgcgagcgcggttcag 816
 Db 1124 CTGAACGGTGAGCTGATCGGTATTAAACACCGGATTTCTGGCGCCGACGCGCAACATC 1183
 QY 817 ggcattctcttgccatcccgatgacgttgccatgaaatgctgcgcaacagctgaaaaac 876
 Db 1184 GGTATCGGCTTCGCTATCCCGATTAACATGTTGTAACCACTGACGTCGCGAGATGTTGAA 1243
 QY 877 accggcaaaagtcacacgcaactggcggtgattattcaggaagatccctacggttg 936
 Db 1244 TACGCCAGGTGAACCGCGGCACTGGGATCATGGGACTGAGCTGAAATCCGGAATTG 1303
 QY 937 gcacagtcgttcgtctggataaaagccagcgcgcatgattgccaatccttcccgcc 996
 Db 1304 GCGAAAGCGATGAAGTCAGCGCCAGCGAGCGCGGTTCGTGAGCAGGTGATGCCGAAT 1363
 QY 997 agccccgcagaaactgcggcctgcaggcggcgacatcgtcctcagcctcagcgcgga 1056
 Db 1364 TCGTCCGCGCGAAGCGGGTATCAAAAGCCGGGATGTATTACCTCGCTGAAACGTAAT 1423
 QY 1057 gaaatagttcttcggcgaccttcccgatcgttcggcgccatcagcgcgggaagaa 1116
 Db 1424 CCGATCAGACGCTTTCGCGCGCTGCGGCTCAGGTGCGCACTATGCCGGTCGCGACAAA 1483
 QY 1117 gtcagctcggcgatggcgcaaaagcggaagaaatcacaatcaaaagccaaagctg 1170
 Db 1484 ATCAGCTCGGCTGCTGCGTGAAGGTAAAGCGATTACGGTGAATCTGGAACGTG 1537
 RESULT 3
 ID T04403 standard; DNA; 1791 BP.
 AC T04403;
 DT 19-AUG-1996 (first entry)
 DE Heat shock protein coding sequence of Rochalimaea henselae.
 KW Rochalimaea henselae; cat scratch disease; bacillary angiomatosis;
 OS CSD; infection; antigen; antibody; vaccine; ds.
 FH Key Location/Qualifiers
 FT cds 141..1652
 FT /*tag= a

[illegible]

Db	1173	GTATTATTTTCGGTAAATGTGAGAAAGATTAAATGATCTCCGTGATCTAGCAAAACGGTATT	1233
Qy	1093	ggcgcaattatcgccgggaaagaagtgcagctcgctggcgatgtgcgcaaggcgagaatac	1152
Db	1233	GCAAAATATGACCCAGAGAAACAGTAACTTAGGAGTTTGGAAATCTGTGTAAGAAGAG	1292
Qy	1153	acaataaagcccaagctgg	1171
Db	1293	AATATTAAAGTTAAACTTG	1311

RESULT 4
 T10423
 ID T10423 standard; DNA; 2894 BP.
 AC T10423;
 DT 03-JUN-1996 (first entry)
 DE H. influenzae SB33 hin47 gene.
 DE Hin47; adhesin; protease; mutagenesis; vaccine; diagnosis;
 KW meningitis; epiglottitis; pneumonia; otitis media; ss.
 OS Haemophilus influenzae strain SB33.
 PH Key Location/Qualifiers
 FT cds 802..2193
 FT /*tag= a
 W09603506-A2.
 PD 08-FEB-1996.
 PF 21-JUL-1995; CA0434.
 PR 21-JUL-1994; US-278091.
 PR 26-AUG-1994; US-296149.
 PR 07-JUN-1995; US-487167.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Chong P, Klein MH, Loomsore SM, Oomen RP, Yang Y;
 DR WPI; 96-117051/12.
 P-PSDB; R32650.
 DR H. influenzae Hin47 protein analogue with decreased protease activity
 PT - and nucleic acid encoding it, useful in vaccines and as diagnostic
 PT agents
 PS Example 2; Fig 2A-2H; 75pp; English.
 CC The hin47 gene (r10423) of H. influenzae SB33 codes for an outer
 CC membrane protein (R92690) thought to be an adhesin conserved
 CC among strains of H. influenzae. The gene was isolated from a
 CC cosmid library of SB33 in lambda EMBL3 using a probe (see
 CC r10424) based on the 5'-end of hin47. Site-directed mutagenesis
 CC of the gene, esp. of codons encoding amino acids 195-201 which
 CC contribute to protease activity, allows prodn. in transformed host
 CC cells of Hin47 analogs useful in vaccines. The gene is also used
 CC in the prodn. of diagnostic probes.
 SO Sequence 2894 BP; 897 A; 570 C; 515 G; 912 T;

Query Match	10.0%;	Score 139.4;	DB 1;	Length 2894;
Best Local Similarity	49.8%;	Prod. NO. 2.3e-28;		
Matches 384;	Conservative	0;	Mismatches 381;	Indels 6; Gaps 1;
382	gccttcacatcagcaaaacggtcacatcctgacacaataccaccagtcggttcgcggtatg	441		
1108	GTCAATTATTTCACGACCAAGGCTATGTTTAAACCAATAATCATCTGTTATTGATGAAGCT	1167		
442	ggcagtatcaaaagtcctgctsaacagacacgcgcgaataataccgcacaaactcatcggttcg	501		
1168	GATAAAATTTACCGTGCGAATTACAAGATGGCGGTGAATTTAAAGCAAAATTTAGTGGGTAAA	1227		
502	gatgccaatccagatgctgcctctctgaaatctgcagcaacgcgaagagctaccocctcgctc	561		
1228	GATGAACATCATGATATTGCATTAGTACAGCTTGAAAAACCAAGTAATTTAAACAGAAATC	1287		
562	aaaatcgcgaatccccaaaaattgaaacgcgcgaatggctgcctccatcgccgcgcgcc	621		
1288	AAATTTTCGTGATTCGCACAAATTTACCGGTAGGCGATTTTCACCTTTCGAATTCGGTAATCCA	1347		
622	ttcggctttgacacacgctgacccgcgcgcatcgttcgcgcacaaagg-----cagaagc	675		
1348	TTTGGTTTAGGTCCTCAACTGTGACATCAGGTATTGTTCTCGAATTCGGTGGCTTCAACAGST	1407		

```
QY 676 ctgcccagaagctacacaccccttcatcaaacaccgaagcttgccatcaatccgggcaat 735
Db 1408 TCTGACAGTGGCCTTATGAAACTATATTCAAACCGATGACGAGTAACCGCGGTAAT 1467
QY 736 tccggggcccgctgttcaacttaaaagagcagctgcggcatcaatcgcaatatac 795
Db 1468 TCGGGTGGACGCTAGTAACCTTAAATGGCACTTATTGGAATTAATACCGCAATTATT 1527
QY 796 agccgcagcgccgattcatcttgccatctcttgccatcccgattgacgttgccatgaat 855
Db 1528 TCTCCAAGCGGTGCAATGCAAGAAATTCCTTTCCGATTTCCAAGTAATCAAGCAAGCAAT 1587
QY 856 gtccgcgaacagctgaaacacacccgcgaagctcaaacgcgggacaactggcgctgattatt 915
Db 1588 TTAGTGCACAAATTTTAGAATTTGGTCAAGTGGCTCGCGGATTTGCTTGGTATTAAAGGT 1647
QY 916 caggaaagtatccatcagctgttggcacagctgttcggtctggtatgataaaagcagcgcgcatg 975
Db 1648 GCGCACTCAATCGTATTTAGCCAAAGCCTTTAATGTAAAGCGGCAACAAGGCGCATTT 1707
QY 976 attgcgaataatcttcccgagcagcccgagaaagctgcggcctgcggcgagggcgagcatc 1035
Db 1708 GTAAGTGAAGTTTACCGAAATCTGCTGCTGAAAGAGCAGGACTTAAAGCGGCGATATT 1767
QY 1036 gtccctcagctcagcgcgagaaatcgttcttccgggacattcccgctcatggtcggc 1095
Db 1768 ATCAGCGGATGACCGTCAAAATAATCTCAAGTTTCTGCTGAAATTCGTGCAAAATCGCA 1827
QY 1096 gccattacgcgggaaagaaagtcagctcggcgctggtatggcgcaagcgaa 1146
Db 1828 ACCATGGTGCAGCAAGAGATTAGCTTGACTTACCTTACGTGATGCGCAA 1878

RESULT 5
X14164
ID X14164 standard; DNA; 1529 BP.
AC X14164;
DE 31-MAR-1999 (first entry)
DE H. pylori GHP0.536 gene.
KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 230..1393
FT /*tag= a
PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.
DR P-PSDB: W98445.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 769-771; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 1529 BP; 497 A; 277 C; 355 G; 400 T;

Query Match 9.2%; Score 128.8; DB 1; Length 1529;
Best Local Similarity 55.0%; Pred. No. 1.3e-25;
```

```
Matches 319; Conservative 0; Mismatches 252; Indels 9; Gaps 3;
QY 373 ttcggttcgggtcttcattcatcagcaaaacggctacatcctgaccaataccccacgtcgtt 432
Db 260 TTAGCAGCGCGGTAAATCATTTCTTAAGAGCGCTATATTGTAACATAACCATGTGATT 319
QY 433 gccggatgggcagatcaaaagt---cctgctcaacgcaaacgcaatataccgccaac 489
Db 320 GATGCGCGGATAAAGATTAAAGTTTACCATTCCAGGAGCAATAAAGAAATATTCCGCCACT 379
QY 490 ctcatcgttctcgatgtccaatcccatgctgcctcttgaataatcgaccaacgcaagag 549
Db 380 CTAGTAGCACCGATCTCTGAAGCGATTTAGCGGTGATTTCGCATCA---CTAAGACAAT 436
QY 550 ctaccctcgtcaaaatcggaatcccaaaatattgaacccgggggaatgggtcgcgtgcc 609
Db 437 CTGCCACGATCAAAATTTCTGATTTCTAATGATATTTCAGTGGCGGATTTGGTTTTGCG 496
QY 610 atcggcgcccttcggttggacaaacggctgacccggcgcatcgtgtccgc---caaa 666
Db 497 ATTGGTAACCCCTTTTGGCGTGGCGAAAGCGTTAGCAAGGATTTGTTTCAGCGCTCAAT 536
QY 667 ggcagaagcctgcgcaacgaaagctacacacccttcacaaacgacgcttgccatcaat 736
Db 557 AAAGCGGGATTGGGATCAACAGCTATGAGAAATTTTCATTCAAACAGACGCTTCCATCAAT 616
QY 727 ccgggcaattccggcgcccgctgttcaacttaaaagacaggtcgctcgcatcaattcg 786
Db 617 CTGGAATTTCCGGCGCGCTTTAATTATGATAGCCGTGGAGGTTAGTGGGATTAATACC 676
QY 787 caatatagccgcagcggggattcatgggcatctccttggccatcccgattgacgtt 846
Db 677 GCTATTATCTCTAAAACTGGGGCAACCAACGCGATTTGGCTTTGCCATCCCTCTTAACATG 736
QY 847 gccatgaatgctgcgcaacagctgaaacacccgcaagtcacaacgcaacactgggc 906
Db 737 GTTAAAGATAGTGTAAACCACTCATCAAAACCGGTAGATTGAAGAGAGGTTACTTGGGC 796
QY 907 gtgattattcagggaagtacctacggtttggcacagtcgt 946
Db 797 GTGGCTGCTCAAGATTGAGTGGCGATTTGCAAAATTTCTT 836

RESULT 6
V90845
ID V90845 standard; DNA; 1540 BP.
AC V90845;
DE 18-FEB-1999 (first entry)
DE Nucleotide sequence of cluster 4.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN WO9849314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
PR 25-APR-1997; US-045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, McAtee CP;
DR WPI: 99-009433/01.
PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 20; Page 282; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC immune response that persists even after antimicrobial treatment. In
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
```


CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma). 411 T;
CC Sequence 1540 BP; 495 A; 280 C; 352 G; 411 T;
SQ

Query Match 8.9%; Score 124; DB 1; Length 1540;
Best Local Similarity 54.5%; Pred. No. 2.7e-24;
Matches 316; Conservative 0; Mismatches 255; Indels 9; Gaps 3;

Qy 373 ttccgttcgggttcctcatcagcaaaacggctacatccctgacccaataccaccgctcgtt 432
Db 300 TTAGCAGCGGTGTCATCATTTCTTAAGACGGCTATATTGTAACTAATAACCATGTGATT 359
Qy 433 gccggtatggcgagtcataaaagt---cctgctcaacgcaagcgaatataccgcca 489
Db 360 GATGCGCGGATTAAGATTAAAGTGACCAATCCAGGAGCAATAAAGAAATATTCGCTACT 419
Qy 490 ctcatcggttcggatgtccaatcccgatgtcgccctctgaaatcgacgcaacggaagag 549
Db 420 TTAGTAGGACCGATCTCGAAGCGATTTAGCCGTGATTTCGCATCA---CTAAGACAAAC 476
Qy 550 ctaccgctgcgaatcgcaatcccaaaatttgaaacccggcgagtgctgcgc---caaa 609
Db 477 TTGCCCACGATCAAAATTCCTGATTTCTAAGCGATTTTACGTTGGCGGATTTGGTTTGGC 536
Qy 610 atcgcgccgcttcggcttttgaacagcgtgacccggcgatcggtgcgc---caaa 666
Db 537 ATTGGTAACCCCTTTTGGCGTGGTGGAAGCGTTACTCAAGGCAATTTTCAGGCGCTCAAT 596
Qy 667 ggcagaagcctgcggcgaagcgtacacaccccttcacacccgagcgtgcaatcaat 726
Db 597 AAAAGCGGATTTGGGATCAACAGCTATAGAAATTTCAATCAACACAGCGCTCTAATAAT 656
Qy 727 ccgggcaatccggcgccgctgttcaactaaaggacagcgtcgctgcgcataatcg 786
Db 657 CTTGGAATTTCCGGCGCGCTTTAATTTAGTAGCGGTGGAGGGTTAGTGGGGATTAATACC 716
Qy 787 caaatatagcgcgagcggcgatcgatggcgatcctcttggcaccgagtgacgtt 846
Db 777 GTTAAAGATATTGTAAACCACTCATCAAAACCGGTGAAGATTGAAGAGGTACTTGGGC 836
Qy 907 gtgattattcaggagtagtattcctacggtttggcacagtcgt 946
Db 837 GTGGGCTTGCAAGATTGAGCGGCGATTTGCAAAATTTCTT 876

RESULT 7
V90921
ID V90921 standard; DNA; 1484 BP.
AC V90921
DT 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone b8.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN W09849314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
PR 25-APR-1997; US-045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, McAtee CP;
WPI: 95-009433/01.
DR New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 37; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive

CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma). 398 T;
CC Sequence 1484 BP; 472 A; 270 C; 344 G; 398 T;
SQ

Query Match 8.7%; Score 120.8; DB 1; Length 1484;
Best Local Similarity 54.1%; Pred. No. 1.9e-23;
Matches 314; Conservative 0; Mismatches 257; Indels 9; Gaps 3;

Qy 373 ttccgttcgggttcctcatcagcaaaacggctacatccctgacccaataccaccgctcgtt 432
Db 299 TTAGCAGCGGTGTCATCATTTCTTAAGACGGCTATATTGTAACTAATAACCATGTGATT 358
Qy 433 gccggtatggcgagtcataaaagt---cctgctcaacgcaagcgaatataccgcca 489
Db 359 GATGCGCGGATTAAGATTAAAGTGACCAATCCAGGAGCAATAAAGAAATATTCGCTACT 418
Qy 490 ctcatcggttcggatgtccaatcccgatgtcgccctctgaaatcgacgcaacggaagag 549
Db 419 TTAGTAGGACCGATCTCGAAGCGATTTAGCCGTGATTTCGCATCA---CTAAGACAAAC 475
Qy 550 ctaccgctgcgaatcgcaatcccaaaatttgaaacccggcgagtgctgcgc---caaa 609
Db 476 TTGCCCACGATCAAAATTCCTGATTTCTAAGCGATTTTACGTTGGCGGATTTGGTTTGGC 535
Qy 610 atcgcgccgcttcggcttttgaacagcgtgacccggcgatcggtgcgc---caaa 666
Db 536 ATTGGTAACCCCTTTTGGCGTGGTGGAAGCGTTACTCAAGGCAATTTTCAGGCGCTCAAT 595
Qy 667 ggcagaagcctgcggcgaagcgtacacaccccttcacacccgagcgtgcaatcaat 726
Db 596 AAAAGCGGATTTGGGATCAACAGCTATAGAAATTTCAATCAACACAGCGCTCTAATAAT 655
Qy 727 ccgggcaatccggcgccgctgttcaactaaaggacagcgtcgctgcgcataatcg 786
Db 656 CTTGGAATTTCCGGCGCGCTTTAATTTAGTAGCGGTGGAGGGTTAGTGGGGATTAATACC 715
Qy 787 caaatatagcgcgagcggcgatcgatggcgatcctcttggcaccgagtgacgtt 846
Db 716 GCTATCATCTCTAAACTGCGGCGCAACCGCATTTGGCTTGGCATCTCTTCTAATCATG 775
Qy 847 gccatgaatgcgcgagcgaagcgtgaaacacccggaagtcacacgagcgaactgggc 906
Db 776 GTTAAAGATATTGTAAACCAATTCATCAAAACCGGTGAAGATTGAAGAGGTACTTGGGC 835
Qy 907 gtgattattcaggagtagtattcctacggtttggcacagtcgt 946
Db 836 GTGGGCTTGCAAGATTGAGCGGCGATTTGCAAAATTTCTT 875

RESULT 8
V31239/C
ID V31239 standard; DNA; 558 BP.
AC V31239
DT 01-OCT-1998 (first entry)
DE E. coli J96 pathogenicity island contig #53.
KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
KW PAI V; pheV; vaccine; protective immune response; ds.
OS Escherichia coli.
PN W09822575-A2.
PD 28-MAY-1998.
PF 21-NOV-1997; U21347.
PR 14-OCT-1997; US-061953.
PR 22-NOV-1996; US-031626.
PA (HUNA-) HUMAN GENOME SCI INC.
PA (UYWI-) UNIV WISCONSIN.


```
OS Rattus sp.
PN WO909046-A1.
PD 25-FEB-1999.
PF 21-AUG-1997; U17297.
PR 21-AUG-1997; US-056453.
PA (KOHN/) KOHN K I.
PA (QUAR-) QUARK BIOTECH INC.
PI Einat P. Skalliter R.
DR WPI; 99-180964/15.
PT New polynucleotides with hypoxia, angiogenesis and apoptosis
PT regulated activity - useful for diagnosis of ischemia, and for
PT treatment involving gene therapy
PS Claim 1: Page 56-57; 72pp; English.
CC Sequences X29126 to X29136 represent isolated polynucleotides with
CC hypoxia, angiogenesis and/or apoptosis regulated activity. The invention
CC provides a method for diagnosing ischemia from bodily fluid samples by
CC identifying at least one expressed gene, or identifying at least one up-
CC regulated gene in a tissue sample. In situ analysis was performed for the
CC above sequences in solid tumours and hypoxic retinas. The expression of
CC the genes is activated in tumours around the hypoxic region in the tumour
CC centre, therefore are hypoxia-regulated in vivo. up regulation promotes
CC angiogenesis required for tumour growth. The genes, proteins,
CC antagonists, antisense oligonucleotides and dominant negative peptides
CC are useful for regulating angiogenesis, apoptosis and hypoxia.
SQ Sequence 607 BP; 157 A; 148 C; 166 G; 134 T;

Query Match 6.1%; Score 85; DB 1; Length 607;
Best Local Similarity 53.9%; Pred. No. 6.6e-14;
Matches 226; Conservative 0; Mismatches 180; Indels 13; Gaps 2;

QY 376 ggtcggcctcatcatcagcaaaacggctacatcctgacaaatccacacgtgtgccc 435
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 GGTCAGGATTCATTGTGTCGGAGGATGGACTGATTGTGCACAAATGCTCAGTGGTGACC 164
QY 436 ggtatggcagtagtaagtcctgctcaacgacgaacgacgaatataccgcgaactcatc 495
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 AACAAAAACCGGTCAGGTTGGCTGAGAAATGAGCGGACTTATGAAGCCAAATCAAG 224
QY 496 ggtcggatgtccaatcgtatgctgcctctgaaatcagcgaacgaagagctacccc 555
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 GATGTGGATGAAAAGGCCGACATTTGCGCTTATCAAGATTGACCACGAGGGTAAAGTGC 284
QY 556 gtctgcaaaatcggaatcccaaaattgaaacggcgcaaatggtcgctgcccacggc 615
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 GTCTGCTGCTTGGCGCTCTCGAGCTCGGCCGACGAGAGAAATTTGTTGGTCCATCGGA 344
QY 616 ggcctcttggtcttgacaacagcgtgacccgctgctgctcgcgaagcagaagc 675
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 AC-CCCTTCTCTTCAAAACACGGTCACCACTGGGATCGTCAGCACACCCACCGGAGGC 403
QY 676 -----ctgcccacgaagatgacacaccccttcatccaaacgacgttgccatc 723
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 GCAAAAGAACTGGGGCTCCGGAACCTCGGATATGGACTACATTTCAGACAGAGCCCATC 463
QY 724 aatccgggcaattccggcgcccgctgttcaacttaaaagacaggtcgctcgccatcaa 782
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 AATTATGGAACCTCCGAGGCCCTTGTAGTAAACCTGGATGGCGAGGTGATTGGGATTAA 522

RESULT 13
V90593/C
ID V90593 standard; DNA; 284 BP.
AC V90593;
DE 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone Y107-173.SEQ from cluster 4.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN WO9849314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
```

```
PR 25-APR-1997; US-045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, McAttee CP;
DR WPI; 99-009433/01.
PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 124; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ Sequence 284 BP; 75 A; 69 C; 60 G; 80 T;

Query Match 5.1%; Score 71; DB 1; Length 284;
Best Local Similarity 56.1%; Pred. No. 3.1e-10;
Matches 134; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 682 aacgaagctacacacccttcacaaacggcagcttgccatcaatccgggcaattccggc 741
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 ATCAACAGCATATGAGAAATTTTCATTCAAACAGACGCGCTCTATTAAATCTGGAATTCGGC 184
QY 742 gcccgcgtgttcaacttaaaaggacagctgcgcggcatcaatccgaaatatacacgcc 801
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 GCGCTTTAAATGTAGTACGCGGTAGTGGGGATTAATACCGCTATCTCTATAA 124
QY 802 agcggcggtatcatgggcatctcttggccatcccgattgacgttgccatgaatgcgcc 861
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 ACTGGGGCAACCCAGCGCATTTGGCTTTGCCATCCCTTCAACATGGTTAAAGATATTGTA 64
QY 862 gaacagctgaaaaaacacccggcaaaagtcacacgcaactggcggtgatttcaggga 920
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 ACCCAACTACTCAAAACCGGTAAAGATTGAAGAGGTTACCTGGCGGTGGCTTGCAGA 5

RESULT 14
V43618
ID V43618 standard; DNA; 1553 BP.
AC V43618;
DE 24-SEP-1998 (first entry)
DE Human secreted protein 18 encoding DNA.
KW Secreted protein; human; cell proliferation; cytokine activity;
KW tissue growth; cellular differentiation; regeneration; activin;
KW inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW anti-inflammatory activity; biomarker; ss.
OS Homo sapiens.
PI Key Location/Qualifiers
FT CDS 65..1495
FT /*tag= a
FT /product= "human secreted protein"
PN WO9825959-A2.
PD 18-JUN-1998.
PF 11-DEC-1997; U22787.
PR 11-DEC-1996; US-032757.
PA (CHIR ) CHIRON CORP.
PI Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR WPI; 98-348453/30.
DR P-PSDB: W63698.
DR Secreted human polypeptides - having cytokine, cell proliferation or
DR differentiation, activin or inhibin, tumour inhibition or
PT anti-inflammatory activities
PS Claim 6; Pages 45-46; 78pp; English.
CC This DNA encodes a human secreted protein. The specification provides
CC secreted protein sequences (W63681 to W63699) encoded by the nucleic
CC acid sequences shown in V43601 to V43619. The invention provides a
```

CC method of identifying a secreted polypeptide which is modified by rough
CC microsomes. The secreted proteins can be used in assays to determine
CC biological activities, such as cytokine, cell proliferation, or cellular
CC differentiation activities, tissue growth or regeneration, activin or
CC inhibin activity, chemotactic or chemokinetic activity, tumour inhibition, or
CC thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC anti-inflammatory activity. The proteins can also be used as biomarkers,
CC to identify tissues or cell types which express the proteins, or a stage-
CC or disease-specific alteration in protein expression. They can be used
CC in protein interaction assays, to identify ligands or binding proteins.
CC Compounds which affect the biological activities of the secreted proteins
CC or their ability to interact with specific ligands can be identified
CC using the proteins in screening assays. The proteins and antibodies that
CC bind specifically to the protein can also be used to design diagnostic
CC tests and therapeutic compositions for diseases which may be associated
CC with altered expression of these proteins. Fusion proteins comprising,
CC e.g. signal sequences or transmembrane domains of the proteins can be
CC used to target other protein domains to cellular membrane or they can
CC be secreted extracellularly.
SQ Sequence 1553 BP; 348 A; 384 C; 467 G; 354 T;

Query Match 4.9%; Score 58; DB 1; Length 1553;
Best Local Similarity 49.5%; Pred. No. 3.7e-09;
Matches 213; Conservative 0; Mismatches 205; Indels 12; Gaps 1;
Qy 368 tgaacttcggttcggttcacatcagcaaaacggtacatcctgacccaataccacg 427
Db 660 TGTACAGTGGCTCTGGGTTCAATAGTCTGTGAGGAGGGCTCATTTATCCCAATGCCCATG 719
Qy 428 tcgttcggtatggcagatcaaaagtcctgctcaacgacaaagcggaatataccgcca 487
Db 720 TTGTCAGGAACACGACAGTGGATTGAGTGGTGTCTCCAGAATGGGGCCGCTTATGAAGCTG 779
Qy 488 aactcatcggttcggtatccaatcgatgtcgcccttcgaaatcgacgcaacggaag 547
Db 780 TTGTCAGGATATTGACCTTAAATTTGGATCTTGGCGGTGATTAGAGATTGAATCAATGCTG 839
Qy 548 agtccaccgtgcacaaatcggcaatcccaaaattgaaacggcggaatgggtcgctg 607
Db 840 AACTTCCTGTACTGATGCTGGGAGATCATCTGACCTTCGGGCTGGAGAGTTTGGTGG 899
Qy 608 ccacggtcgcccttcggttcgttgaacacagcgtgacccggtcggtcggtcgccaaag 567
Db 900 CTTTGGCAGGCCATTTTCTCTCGACACACACAGCTACTCGAGGAATTCAGCAGCCAAAC 959
Qy 668 gcagaag-----cctgcccaacgaaagctacacaccccttcacccaacccgacg 715
Db 960 AGCGAGGGGGCAAGAACTGGGGATGAAGATTACAGATATGGACTACGTCCAGATTGATG 1019
Qy 716 ttgccaatcaatccgggaattcggcgcccgctgttcactaaagacaggtcgctcg 775
Db 1020 CCACAAATTAACATGGAATTTCTGTGGTGGTCTCTGTGTAATGATGGATGGTGTGATG 1079
Qy 776 gcatcaattc 785
Db 1080 GCGTCAATTC 1089

RESULT 15
V90690
ID V90690 standard; DNA; 273 BP.
AC V90690;
DT 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone Y261T3.SEQ from cluster 4.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN W09849314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
PR 25-APR-1997; US-045107.

PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, Mcatee CP;
DR WPI; 99-009433/01.
PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 155; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ Sequence 273 BP; 78 A; 58 C; 64 G; 73 T;

Query Match 4.5%; Score 62.8; DB 1; Length 273;
Best Local Similarity 54.9%; Pred. No. 5e-08;
Matches 124; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Qy 692 aacgaagactacacacccttcacccaacacgacgttgccatcaatccggcaattccgac 741
Db 48 ATCAACAGCTATGAGATTTCATTCAAAACAGACGCTCTATTAATCTCTGGAATTCGCGC 107
Qy 742 gcccgcgtgtcaacttaaaagacagtcgtcgccatcaatccgcaatatacacacgcgc 801
Db 108 GCGGCTTTAATTGATAGCGGTGAGGGTTAGTGGGGATTATACCGCTATCATCTCTAAA 167
Qy 802 agcggcggtatccatgggcatccttggccatcccgatgacggtgcccataatgctgcc 861
Db 168 ACTGGGGCAACACCGGCATTTGGCTTTGCCATCCCTCTTAACATGTTAAAGATATTGTA 227
Qy 862 gaacagctgaaaaaacacacggcaaaagtccaacgacgacactggcg 907
Db 228 ACCCAACTCATCAAAACCGGTTAGATTGAAGAAGTTACTTGGGCG 273

Search completed: April 14, 2000, 16:52:18
Job time: 6877 sec

Tue Apr 18 14:25:00 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:28:42 ; Search time 83.49 Seconds
(without alignments)
2000.405 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgtcaaaaataccaata.....tcagggaagtcctccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues
Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCUT9_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175.6	12.6	1980	2	US-08-350-741-1
2	175.6	12.6	1980	4	US-08-463-875A-1
3	154.2	11.1	1791	1	US-08-245-294-7
4	154.2	11.1	1791	2	US-08-474-499-7
5	154.2	11.1	1791	2	US-08-307-279A-7
6	154.2	11.1	1791	6	PCT-US95-06211-7
7	139.4	10.0	2894	1	US-08-278-091-1
8	139.4	10.0	2894	1	US-08-483-859-1
9	139.4	10.0	2894	1	US-08-472-173-1
10	139.4	10.0	2894	3	US-08-487-167-1
11	139.4	10.0	2894	4	US-08-482-816-1
12	139.4	10.0	2894	4	US-08-296-149-1
13	139.4	10.0	2894	4	US-08-801-499-1
14	139.4	10.0	2894	4	US-08-615-271-1
15	136.2	9.8	1611	1	US-08-485-569-1
16	136.2	9.8	1611	1	US-08-480-993-1
17	136.2	9.8	1611	3	US-07-903-079B-1
18	85.8	6.2	539	1	US-08-322-742-18
19	41	2.9	1187	2	US-08-440-856A-2
20	39.2	2.8	2414	7	5248599-1
21	37.2	2.7	2793	2	US-08-209-747-1
22	37.2	2.7	2793	2	US-08-458-298-1
23	36.8	2.6	716	4	US-08-211-718-1
24	35.8	2.6	1608	4	US-08-211-718-8
25	35.8	2.6	2167	3	US-08-461-775-9
26	35.8	2.6	2668	3	US-08-461-775-11
27	35.6	2.6	424	4	US-08-476-176B-7

Sequence 2, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 10, Appli
Sequence 10, Appli
Sequence 185, App
Sequence 185, App
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 1, Appli

28 35.6 2.6 1560 4 US-08-709-979A-2
29 35.4 2.5 783 1 US-08-264-861A-11
30 35.4 2.5 783 6 PCT-US95-07784-11
31 35.4 2.5 1150 1 US-08-264-861A-10
32 35.4 2.5 1150 6 PCT-US95-07784-10
33 35.2 2.5 898 4 US-08-997-080-185
34 35.2 2.5 898 4 US-08-997-362-185
35 34.8 2.5 28958 1 US-08-456-837-6
36 34.8 2.5 28958 1 US-08-456-837-6
37 34.8 2.5 28958 1 US-08-457-342-6
38 34.8 2.5 28958 2 US-08-457-342-6
39 34.8 2.5 28958 2 US-08-458-076A-6
40 34.8 2.5 28958 2 US-08-764-233A-4
41 34.8 2.5 28958 2 US-08-457-335A-6
42 34.8 2.5 28958 2 US-08-729-214-6
43 34.8 2.5 49377 2 US-08-764-233A-1
44 34.6 2.5 833 3 US-08-403-852D-3
45 34.6 2.5 5392 3 US-08-403-852D-1

ALIGNMENTS

RESULT 1
US-08-350-741-1
; Sequence 1, Application US/08350741
; Patent No. 5804194
; GENERAL INFORMATION:
; APPLICANT: DOUGAN G.,
; APPLICANT: CHARLES I.G.,
; APPLICANT: HORMACHE C.E.,
; APPLICANT: JOHNSON K.S.,
; APPLICANT: CHATFIELD S.N.,
; TITLE OF INVENTION: LIVE VACCINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON and VANDERHYE PC
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/350,741
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 30-NOV-1992
; APPLICATION NUMBER: GB 9007194.5
; FILING DATE: 30-MAR-1990
; APPLICATION NUMBER: PCT/GB91/00484
; FILING DATE: 28-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double


```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1652
US-08-307-279A-7
```

```

Query Match      11.1%; Score 154.2; DB 2; Length 1791;
Best Local Similarity 50.4%; Pred. No. 1.5e-33;
Matches 403; Conservative 0; Mismatches 393; Indels 3; Gaps 1;

QY 373 ttcgggttcgggttcacatcagcaaaaacggctcatcctcctgacccaataccacgcgtt 432
DB 516 TTTGGATCGGGTTTTTTTATCTCGTCTGATGTTATTTGTGACCAATAATCATGTGATT 575

QY 433 gccgggtatggcgagatcaaaagtcctcgtcaacgcaagcggaataataccgccaaactc 492
DB 576 TCTGATGGCACAAGTTACGCTGTTTCTTGTGATGCGGTACAGAACTGAATGCAAAACTC 635

QY 493 atcgggttcgggttcacatcagcaaaaacggctcatcctcctgacccaataccacgcgtt 552
DB 636 ATTGGAACGGACCCACGACTGATCTTGAAGTCAATGAAAAAGAAAAATTT 695

QY 553 cccggtcgtcaaaaatcgcccaatcccccaaaatttgaaacggggcggaatgggtcgtccatc 612
DB 696 TCGTACGTTGATTTTGGTGATGATTCAAAACCTTCGTTGGTGATTTGGTGTTCGTGATT 755

QY 613 ggcgcgccttcggtttgacaaagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 672
DB 756 GGTAAATCCATTTGGTCTGTTGGTGAACCTGTGACAGAGGTTGTTTCAGCAGCTGGACGT 815

QY 673 agcctgcccacgaagctcacacaccttcacccaacggcgagcttgcccatcaatccgggc 732
DB 816 GATATCGGTACCGGTTTATGATGATTTTATTCAGATTGATGCTCAGTTAATCGAGGA 875

QY 733 aattccggcgccgcgtgttcaacttaaaagacaggttcggtcgcgcgcgcgcgcgcgcgcgcgc 792
DB 876 AATTCTGGAGGTCCAACTTTTGTATCTTAACGGAAAGGTTGTTGGAGTGAATACGGCAATT 935

QY 793 tacagccgcagcgcggtattcgtggtcgtcctcttgcctcccgatgacgttgccatg 852
DB 936 TTTTCTCTCTTCGTTGGGGCAACGTTGGGATGCTTTTCGCTATTCGGCAGCAACAGCGAAC 995

QY 853 aatgtcgcgaacagctgaaaaacacccggcaaaagtcctcctcctcctcctcctcctcctcct 912
DB 996 GAGGTTGTGCAACAACCTTATCGAAAAAGGTTTATTCAGCGTGGTTGGGTTTCAG 1055

QY 913 attcagaagatcctcactcgtttggtgacagctcgttcggtcgtgataaagccagcgcgca 972
DB 1056 ATTACGCTGTAAACAAAAGAAATTTCTGATTCATTTGGTTTGAAGGAGGCTAAGGTGCG 1115

QY 973 ttgattgccaataatcctcccgagcccgagcccgagacgctgcggcctgcaggcgggcgac 1032
DB 1116 TTAATTACCGATCCATTAAAGGGCCAGCGCGCAAAA--GCTGGTATCAAGGACGAGTAT 1172

QY 1033 atcgtcctcagctcagcgcgggagaaatacgtttcttcggcgacacctcccgctcagtc 1092
DB 1173 GTTATTATTTTCGGTAAATGTTGAGAGATTAAATGATGTCGCTGATCTAGCAAAAGCGTATT 1232

QY 1093 ggcgcctattacgcccgaagaaagatcagctcgcgcgtatggcgcaaaagcgaagaaatc 1152
DB 1233 GCAAAATATGAGCCAGGAGAAACAGTAACCTTAGGAGTTTGGAAATCTGTTAAAGAGAG 1292

QY 1153 acaatcaaaagccaagctgg 1171
DB 1293 AATATTAAGGTTAACTTG 1311
```

RESULT 6

PCT-US95-06211-7

; Sequence 7, Application PC/TUS9506211

; GENERAL INFORMATION:

```

; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06211
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18 MAY 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.6121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1649
PCT-US95-06211-7
```

```

Query Match      11.1%; Score 154.2; DB 6; Length 1791;
Best Local Similarity 50.4%; Pred. No. 1.5e-33;
Matches 403; Conservative 0; Mismatches 393; Indels 3; Gaps 1;

QY 373 ttcgggttcgggttcacatcagcaaaaacggctcatcctcctgacccaataccacgcgtt 432
DB 516 TTTGGATCGGGTTTTTTTATCTCGTCTGATGTTATTTGTGACCAATAATCATGTGATT 575

QY 433 gccgggtatggcgagatcaaaagtcctcgtcaacgcaagcggaataataccgccaaactc 492
DB 576 TCTGATGGCACAAGTTACGCTGTTTCTTGTGATGCGGTACAGAACTGAATGCAAAACTC 635

QY 493 atcgggttcgggttcacatcagcaaaaacggctcatcctcctgacccaataccacgcgtt 552
DB 636 ATTGGAACGGACCCACGACTGATCTTGAAGTCAATGAAAAAGAAAAATTT 695

QY 553 cccggtcgtcaaaaatcgcccaatcccccaaaatttgaaacggggcggaatgggtcgtccatc 612
DB 696 TCGTACGTTGATTTTGGTGATGATTCAAAACCTTCGTTGGTGATTTGGTGTTCGTGATT 755

QY 613 ggcgcgccttcggtttgacaaagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 672
DB 756 GGTAAATCCATTTGGTCTTGGTGAACCTGTGACAGAGGTTATCTTTCAGTTCAGTTCAGGAG 815

QY 673 agcctgcccacgaagctcacacaccttcacccaacggcgagcttgcccatcaatccgggc 732
DB 816 GATATCGGTACCGGTTTATGATGATTTTATTCAGATTGATGCTCAGTTAATCGAGGA 875

QY 733 aattccggcgccgcgtgttcaacttaaaagacaggttcggtcgcgcgcgcgcgcgcgcgcgcgc 792
DB 876 AATTCTGGAGGTCCAACTTTTGTATCTTAACGGAAAGGTTGTTGGAGTGAATACGGCAATT 935
```

QY 793 tacagcgacgagcggttcatggtggaatctctcttggccatcccgattgacgttcccatg 852
Db 936 TTTTCTCTCTCTGGGGCAACGTTGGGATGCTTTTCGTATTCCGGCAGCAACAGCGAAC 995
QY 853 aatgcgcgacagctgaaacacccgcaaaagtcacacgcaacgcaactggcgctgatt 912
Db 996 GAGGTGTGCAACAACTTATCGAAAGGTTTGTAGTTCACGCTGGTGGCTTGGGGTTCAG 1055
QY 913 attcagaagtaactcactacggtttggcacagtcgtcttcggtctgtgataaaagccgcgca 972
Db 1056 ATTGAGCTGTAAACAAAGAAATTTCTGATTCATTAATTTGTTTGAAGGAGGCTAAAGTGCG 1115
QY 973 ttgattgccaaaactctcccgagcccgagcccgagcaactgcccgcctgcagcgggcgac 1032
Db 1116 TTAATTACCGATCCATTAAAGGGGCGACCGCAAAA--GCTGGTATCAAGGCGAGTGAT 1172
QY 1033 atcgtctcagctcgcagcgcggaatacgttcttcggcgagaccttcccgatggtc 1092
Db 1173 GTTATTAATTCGGTAATGGTGAGAGATTAATGATGTCGGTGATCTAGCAAGCGTATT 1232
QY 1093 ggcgccattcgcggggaagagagtcgctcgcggtatggcgcaaaagcgcaaaatc 1152
Db 1233 GCAATATGAGCCCGAGGAGAAACAGTAACTTAGGAGTTTGGAAATCTGGTAAAGAGAG 1292
QY 1153 acaatcaagcgaagctgg 1171
Db 1293 AATATTAAGGTAAACTTG 1311

RESULT 7

US-08-278-091-1
; Sequence 1, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278.091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-278-091-1

Query Match 10.0%; Score 139.4; DB 1; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;
QY 382 gcttcatcatcagcaaaacggctacatctctgaccaataccccacgctcgttgcgcgtatg 441
Db 1108 GTCATTTAATGCAAGCAAGGCTATGTTTAAACCAATATCATGTTATGATGAAGCT 1167
QY 442 ggcagtatcaagtcctgctcaacgacaagcgcaataataccgcaaacatcatcggttcg 501
Db 1168 GATAAAATTAACGTCGAATTTACAGATGGCGTGAATTTTAAAGCAAAATTTAGTGGGTAAA 1227
QY 502 gatgtccaatccgatctcgcctcttgaataatcgacgcaacggaagagctaccctgcgtc 561
Db 1228 GATGAACATATCAGATATTGCATTAGTACAGCTTGAAAAACCAAGTAATTTAAACAGAAATC 1287
QY 562 aaaaatcgcaatcccaaaaatttgaaacccggcggaatggctcgtcgcacatcgcgcgccc 621
Db 1288 AAATTTGCTGATTCGACAAATTTACGCTAGGCGATTTTCACTGTTGCAATCGGTAATCCA 1347
QY 622 ttcggctttgacaacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 675
Db 1348 TTTGGTTTAGTCAAACTGTGACATCAGGTATTGTTCTGCATTGGTCTGCTTCAACAGT 1407
QY 676 ctgcccacgaaagctacacaccccttcacccaaacgcaagcttgcacatcaatccgggcaat 735
Db 1408 TCTGACAGTGGCACTTATGAAAACTATATTCAAAACCGATCAGCAGTAAACCCGGGTAAT 1467
QY 736 tccggcgccgc 795
Db 1468 TCGGGTGGAGCGTTAGTAACTTAAATGGCGAACTTATTGGAATTAATACCGCAATATT 1527
QY 796 agccgcagcgcggtattcattggcgcctctcttggccatcccgattgacgttgccatgaat 855
Db 1528 TCTCCNAGCGGTGGCAATCGAGGAATGCTTTGCCATTCCAAGTAATCAAGCAAGCAAT 1587
QY 856 gtcgcggaacagctgaaacacccggcaaacgctcccaacgcggaactggcggtgattatt 915
Db 1588 TTAGTCAACAAATTTAGAAATTTGGTCAAGTGCCTCGCGGATTTGCTTGGTATTAAAGGT 1647
QY 916 caggaagtatactacggtttggcacagtgctcgtgcttgatgataaagccagcgcgatg 975
Db 1648 GCGCAACTCAATGCTGATTAGCCAAAGCGCTTTAATGTAAAGCGCAACAAAGCGCATTT 1707
QY 976 attgccaaaatctctcccgccagcccgccgcaagcgtgcgcgcgcgcgcgcgcgcgcgcgc 1035
Db 1708 GTAAGTGAAGTTTACCGAAATCTGCTGTGAAAAAGCAGGACTTAAAGCGGGCGATATT 1767
QY 1036 gtcctcagcctcagcgcggaagaaatcacgttcttccggcgacattcccgctcagtggtcgc 1095
Db 1768 ATCAGCGCGATGAACGGTCAAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAAATCGCA 1827
QY 1096 gccattacccgggaagaaagtcacccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1146
Db 1828 ACCACTGGTGCAGGCAAGAGATTAGCTTACTTACTTACTTACTTACTTACTTACTTACTTACT 1878

RESULT 8
US-08-483-859-1
; Sequence 1, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

QY	856	gtcgcgcgaacagctgaaacacacccggcaaaagtcacacgcgcgcgaacatcggcgctgataatt	915
Db	1588	TTAGTGCACAAATTTAGAAATTTGGTCAAGTCGGCTCGCGATGCTTGTGTATTAAAGGT	1647
QY	916	caggaaagtatcctacggtttggcgacagtcgcttcggtctgataaagccagcgcgcgcattg	975
Db	1648	GGCGGAATCAATGCTGATTTAGCCAAAGCCCTTTAATGTAAACGGCGCAACAGGGCGATTT	1707
QY	976	attgcgcaaatcctctccggcgagcccccagaaacgctgcggcctcgcaggcggcgacatc	1035
Db	1708	GTAATGATGATTTACCGAAATCTGCTGCTGAAAAAGCAGGACTTAAAGCGGGCGATATT	1767
QY	1036	gtcctcagcctcgcacgcggcggaataagctcttcgcgcacctcccgctcatggtcgcgc	1095
Db	1768	ATCACGGCGATGACCGGTCAAAAAATCTCAAGTTTCGCTGAANTCTGTCGCAAAATCGCA	1827
QY	1096	gcattacgcgcgggaaagaagtcagcctcgcgcgtatggcgcaaggcgaa	1146
Db	1828	AGCACTGCTGTCAGGCAACGAGATTAGCTTTGACTTACTTACGTGATGGCAAA	1878

RESULT 9
US-08-472-173-1
; Sequence 1, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,173
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-493 MIS.vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-472-173-1

```
Query Match
10.0%; Score 139.4; DB 1; Length 2894
```

ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,859
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-483-859-1

	Query Match	10.08;	Score 139.4;	DB 1;	Length 2894;
	Best Local Similarity	49.89;	Pred. No. 2.3e-29;		
	Matches 384;	Conservative 0;	Mismatches 381;	Indels 6;	Gaps 1;
QY	382	ggcttcacatcagcaaaaacggtacatcctgaccatacccacgtcggtgacggtatg	441		
DB	1108	GTCAATTATTAAAGCAAGCAAGGCTATCTTTTAACCAATAATCATGTATTGATGAAGCT	1167		
QY	442	ggcagatctcaagtcctcctaacaagacgacgaatataccgcacaaactcatcggttcg	501		
DB	1168	GATAAAATTTACCGTCGAATTAACAATGGCGGTGAATTTAAAGCAAAATTTAGTGGGTAA	1227		
QY	502	gatgtcccaatccgtagtgcgcctctgaaaatcgacgcacgaaggaagatccacctcgctc	561		
DB	1228	GATGAACATATCAGATATTGCATTAGTACAGCTTGAAAAACCAAGTAATTTAACAGAAATC	1287		
QY	562	aaaatcgccaatcccaaaaatttgaaacggcgcaatgggtcgctgcctacggcgcgccc	621		
DB	1288	AAATTTGCTGATTCGCACAAATTAACCGTGAAGCGATTTCACTGTGCAATCGGTATCCA	1347		
QY	622	tcgcgctttgacaacagcgtgaccgcgcgcacgtgcgcgcacaaagg-----cagaagc	675		
DB	1348	TTTGGTTTAGGTCMAACTGTGACATCAGGTATTGTTTCTGCAATTTGGGTGCTTCAACAGGT	1407		
QY	676	ctgcgcaaacgaagtacacaccccttcacccaaacgcacgctgcccataatcggcgcaat	735		
DB	1408	TCGTGACAGTGGCACTTATGAAACATATTTCAAACCGATGTCAGCAGTAAACCGCGTAA	1467		
QY	736	tcgcgcgcgcgcgtgttcacattaaaaggacaggtcgtgcggcatcattcgcgaataac	795		
DB	1468	TCGGGTGGACGCTTACTAAACTTTAAATGGCGAACTTATTGGAATTAATACCGCAATATT	1527		
QY	796	agcgcgacgcgcggaattcatgggcatctctcttgccatcccgattacggtgccaatgaat	855		
DB	1528	TCFCCAAAGCGGTGGCAATGCAGGAATTCCTTTGCGATTCCAAAGTATCAAGCAAGCAAT	1587		

```
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggcttcacatcagcaaaacggcgctacatcctgaccataaccacgctgttgcgggtatg 441
Db 1108 GTCAATTAATGCAAGCAAGGCTATGTTTAAACCAATATCATGTTATTGATGAAGCT 1167
QY 442 ggcagtatcaaaactcgtctcaacagcaagcgcaaatatccgcaaaactcatcggttcg 501
Db 1168 GATAAAATTAACCGTGCATTAACAAGATGGCGGTGAATTTAAAGCAAAATTAAGTGGGTAAA 1227
QY 502 gatgtcaatccgatgcgcctctgaaatcgacgaacgcaagagctaccgctcgc 561
Db 1228 GATGAACTATCAGATATGCAATAGTACAGCTTGAACCAACCAAGTAATTTAACAGAAATC 1287
QY 562 aaatgggcaatcccaaaaaatttgaacccggcgcaatgggtcgtgcctcggcgcgccc 621
Db 1288 AAATTTGCTGATTCGACAAATTAACGGTAGCGGATTTCACTGTTGCAATCGGTAATCCA 1347
QY 622 ttcgggtttgacacagcggtgacccggcgcatcgtgtcgcgcaaaagg-----cagaagc 675
Db 1348 TTTGGTTTAGGTGCAAACTGTGACATCAGGTATTTGTTCTGCAATGGGTTCGAACAGGT 1407
QY 676 ctgcccacgaaagctacacaccccttcaaaacggcgcaatgggtcgtgcctcggcgcgccc 621
Db 1408 TCTGACAGTGGCAGCTTATGAAAACCTATATTCAAACCGATGCAGCAGTAACCCGCGTAAT 1467
QY 736 tccggcgcccgctgttcaacttaaaagacaggtcgtcggcatcaattcgcgaataatc 795
Db 1468 TCGGTTGGAGCGTGTAGTAACTTAAATGCGAATTTATTGGAATTAATACCGCAATATT 1527
QY 796 agccgacgagcggttcatgggcatctcttgcctcccgatcccgatcgatcgatcgat 855
Db 1528 TCTCAAGCGGTGGCAATGCAGGAATTTGCTTTTGGGATTTCCAAAGTAATCAAGCAAGCAAT 1587
QY 856 gtcgcccgaagctgaaaaaacccggcaaaagtcacacgagcgacactggcggtgattatt 915
Db 1588 TTAGTGCAACAAATTTAGAAATTTGGTCAAGTGGTGGCGGATTTGCTTGGTATTAAAGGT 1647
QY 916 caggaaagatacctacggttttgacagctcgttcggtcgttgataaagcagcgcgcatg 975
Db 1648 GCGCAACTCAATGCTGATTTAGCAAAAGCCCTTTAATGTAAAGCGGCAACAAAGCGCAATT 1707
QY 976 atgcccataatcttcccgagcccgagacgtgcccgcgtgcgagcggtgagcgagacatc 1035
Db 1708 GTAAAGTGAAGTTTACCAGAAATCTGCTGCTGAAAGAACGAGACTTAAAGCGGCGGATATT 1767
QY 1036 gtccctcagctcagcgcgagaaatacgttcttccggcgacccctcccgctcatggtcgc 1095
Db 1768 ATCAGCGGATGAACGCTCAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAATCGCA 1827
QY 1096 gccattacgcccggaaagaaagtcagcctcggcgatagggcgcaagcgaa 1146
Db 1828 ACCACTGGTGCAGCAAGAGATTAGCTTGACTTACTTACGTACGTGATGGCAAA 1878

RESULT 10
US-08-487-167-1
; Sequence 1, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-167-1

Query Match 10.0%; Score 139.4; DB 3; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggcttcacatcagcaaaacggcgctacatcctgaccataaccacgctgttgcgggtatg 441
Db 1108 GTCAATTAATGCAAGCAAGGCTATGTTTAAACCAATATCATGTTATTGATGAAGCT 1167
QY 442 ggcagtatcaaaactcgtctcaacagcaagcgcaaatatccgcaaaactcatcggttcg 501
Db 1168 GATAAAATTAACCGTGCATTAACAAGATGGCGGTGAATTTAAAGCAAAATTAAGTGGGTAAA 1227
QY 502 gatgtccaaatccgatgcgcctctgaaatcgacgaacgcaagagctaccgctcgc 561
Db 1228 GATGAACTATCAGATATGCAATAGTACAGCTTGAACCAACCAAGTAATTTAACAGAAATC 1287
QY 562 aaatgggcaatcccaaaaaatttgaacccggcgcaatgggtcgtgcctcggcgcgccc 621
Db 1288 AAATTTGCTGATTCGACAAATTAACGGTAGCGGATTTCACTGTTGCAATCGGTAATCCA 1347
QY 622 ttcgggtttgacacagcggtgacccggcgcatcgtgtcgcgcaaaagg-----cagaagc 675
Db 1348 TTTGGTTTAGGTGCAAACTGTGACATCAGGTATTTGTTCTGCAATGGGTTCGAACAGGT 1407
QY 676 ctgcccacgaaagctacacaccccttcaaaacggcgcaatgggtcgtgcctcggcgcgccc 735
Db 1408 TCTGACAGTGGCAGCTTATGAAAACCTATATTCAAACCGATGCAGCAGTAACCCGCGTAAT 1467
QY 736 tccggcgcccgctgttcaacttaaaagacaggtcgtcggcatcaattcgcgaataatc 795
Db 1468 TCGGTTGGAGCGTGTAGTAACTTAAATGCGAATTTATTGGAATTAATACCGCAATATT 1527
QY 796 agccgacgagcggttcatgggcatctcttgcctcccgatcccgatcgatcgatcgat 855
Db 1528 TCTCAAGCGGTGGCAATGCAGGAATTTGCTTTTGGGATTTCCAAAGTAATCAAGCAAGCAAT 1587
QY 856 gtcgcccgaagctgaaaaaacccggcaaaagtcacacgagcgacactggcggtgattatt 915
Db 1588 TTAGTGCAACAAATTTAGAAATTTGGTCAAGTGGTGGCGGATTTGCTTGGTATTAAAGGT 1647
```


;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/296,149
;; FILING DATE:
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stewart, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-390
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2894 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;;
US-08-296-149-1

Query Match 10.0%; Score 139.4; DB 4; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggcttcacatcgcaaaacggctacatcctgacaaatcccaacacgctggtccggtatg 441
Db 1108 GTCAATTATTAAATCAAGCAAGGCTATGTTTAAACCAATATCATGTATTGATGAAGCT 1167

QY 442 ggcagatcaaaagtcctgtcctcaacgcaagcgcaatataccgccaaactcattggttcg 501
Db 1168 GATAAAATTTACCGTGCATTTACAGATGGCGGTGAATTTAAAGCAAAATAGTGGGTAAA 1227

QY 502 gatgtccaatcgatcgccctctgaaatcgagcaagcaagcaagagctaccgtgctc 561
Db 1228 GATGAACATACAGATATGCAATTTAGTACGCTTGAAACCAAGTAATTTAAACAGAAATC 1287

QY 562 aaaaatcggaatcccaaaatgaaacccgagcaatgggtcgctgcacgcgcgcgcgc 621
Db 1288 AAATTTGCTGATTCGCACAAATACCGTAGCGGATTTCACTGTGCAATCGGTAAATCCA 1347

QY 622 ttccggtttgaaacagcggtgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 675
Db 1348 TTTGGTTTATAGTCAAACTGTGACATCAGGTATTTGTTTCTGCAATGGGTCTTCAACAGGT 1407

QY 676 ctgcccacgaaagctacacaccccttcacaaacgagcgttgccatcaatccgggcaat 735
Db 1408 TCTGACAGTGGCACTTATGNAACATATATTCAAACCGATCGCAGGATAAACCGCGGTAA 1467

QY 736 tccggcgcccgctgttcaacttaaaaggacaggtcgctgcgcacatccgcaaatatc 795
Db 1468 TCGGGTGGAGCGTGTAACTTAAATGGCGAATTTATTTGGAATTTATCCGCAATTTATT 1527

QY 796 agccgcagcggtgattcatgggcacatcctcttgcaatcccgatcgctgcaatgaat 855
Db 1528 TCTCCAAGCGGTGGCAATCGAGAAATTTGCCTTTTCGCAATTCGAATCAAGCAAGCAAT 1587

QY 856 gtccgcgaacagctgaaacacccgcaaaagtcacacgagcagcaactggcggtgattatt 915
Db 1588 TTAGTGCACAAATTTTAGNATTTGTCAGTGGCGGCGGATGCTTGTGTTATTAAGGT 1647

QY 916 cagggaagtacctacggttttgacagtggttgcgttggtgataaagccagcgcgcatg 975
Db 1648 GCGCAACTCAATGCTGATTTAGCAAAAGCCTTTAAATGAAGCGCGCAACAAAGGCGCATTT 1707

QY 976 attgcaaaatccttcccgccagcccgcaagacgctgcccgcctgagggcgcgacatc 1035
Db 1708 GTAAGTGAAGTTTACCGAAATCTGCTGCTGAAAGAGCAGGACTTAAAGCGGCGATATT 1767

QY 1036 gtccctcagctcgacggcgagaaatcacgttcttccggcgacaccttccctcgtcggtcgc 1095

Db 1768 ATCAGGCGATGACGGTCAAAAAATCTCAAGTTTCGCTGAATTCGTCAAAAAATCGCA 1827
QY 1096 gccattacgcgggaaagaagtcagctcgctcgctgctgctgctgctgctgctgctgctg 1146
Db 1828 ACCACTGGTGCAGGCAAGAGATTAGCTTACTTACTTACGTGATGGCAAA 1878

RESULT 13

US-08-801-499-1
; Sequence 1, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease A
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801.499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-801-499-1

Query Match 10.0%; Score 139.4; DB 4; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggcttcacatcgcaaaacggctacatcctgacaaatcccaacacgctggtccggtatg 441
Db 1108 GTCAATTATTAAATCAAGCAAGGCTATGTTTAAACCAATATCATGTATTGATGAAGCT 1167
QY 442 ggcagatcaaaagtcctgtcctcaacgcaagcgcaatataccgccaaactcattggttcg 501
Db 1168 GATAAAATTTACCGTGCATTTCAAGATGGCGGTGAATTTAAAGCAAAATTAGTGGGTAAA 1227


```
Qy 502 gatctcaaatccggtatgctgcctctctgaaatacgaacgaaggaagactaccgctgc 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GATGAACATATCAGATATTCGATATAGTACAGCTTGAATAACCAAGTAATTTAACAGAAATC 1287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 562 aaatccggaatcccaaaatttgaacacggcggaatgggtcgctgccatcgcgcgccc 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 AAATTTGCTGATTCGACAAATTAGCGGTAGCGGATTCATCTGTTCGATCGGTAATCCA 1347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 622 ttcggctttgaaacagcgtgacccggcgatcgctgtccgcgaag-----cagaagc 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 TTTGGTTAGGTCAAACTGTGACATCAGGTATGTTTCTGCAATGGGTCAACAGGT 1407
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 676 ctgcccacgaagactacacaccttcacaaacggagcttgccatcaatccgggaat 735
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 TCTGACAGTGGCACTTAGAAACTATATCAAAACCGATGAGCAGTAACCCGGGTAAAT 1467
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 736 tccgggcccgcgtgttcaacttaaaagacaggctcgctcggaataattcgcaaatatc 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 TCGGTGGAGCGTTAGTAAACTTAAATGCGCACTTATTGGAAATTAATACCGCAATATT 1527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 796 agccgacggcggaattcatgggcatctccttggccatcccgatggagcttgccatgaat 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1528 TCTCAAGCGGTGGCAATTCGAGGAATTCCTTTGGGATTCGAAGTAATCAAGCAAGCAAT 1587
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 856 gtcgccgaagcgtgaaacacacggcgaaagtcgaacgcgacacactggcggtattatt 915
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1588 TTAGTGCAACAAATTTAGAAATTTGGTCAAGTGCCTCGCGGATTCCTTGGTATTAAAGGT 1647
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 916 caggaagtactcactcgggtggcacagtcggttcggtcgtggataaagcagcgcgcatg 975
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1648 GCGCAACTCAATGCTGATTTAGCCAAAGCCTTTAAATGAAGCGCGCAACAGGCGCAATT 1707
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 976 attgccaatactctcccgagcccccagaaactgcggcgctcagcgcgcgacatc 1035
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1708 GTAAGTGAAGTTTACCAGAAATCTGCTGCTGAAAAAGCAGGACTTAAAAAGCGGCGATATT 1767
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1036 gtcctcagcctcgacgcgcggaatactgtcttcctcgcgaccttcctcgatgctgcgc 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1768 ATCAGCGCGATGACGGTCAAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAAATCGCA 1827
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1096 gccattacgccccggaagaagtcagctcgcgctatggcgcaaaagcgaa 1146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1828 ACCACTGGTGAGGCAAAAGAGATTAGCTTGACTTACTTACGGTGGGCAAA 1878
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

```
US-08-615-271-1
; Sequence 1, Application US/08615271
; Patent No. 5981503
;
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
;
; CORRESPONDENCE ADDRESS:
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
```

```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-615-271-1

Query Match 10.0%; Score 139.4; DB 4; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

Qy 382 ggcttcacatcatcagcaaaacggctacatcctcgaccaataccccacgtgctgctgcgtatg 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 GTCATTATTATGCAAGCAAGGCTATGTTTAAACCAATAATCATGTTATTGATGAAGCT 1167
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 442 ggcgagtcacaaagtcctgctcaacgacgaagcggaataataccgcaaaactcactcggttcg 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 GATAAAATTACCGTGCATTTACAAAGATGGCGGTGAATTTAAAGCAAAATTTAGTGGGTAAA 1227
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 502 gatgccaatccgatgctgcctcttgaaatacgaacgaacgaagagctaccgctcctc 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GATGAACATCAGATATTGTCATTTAGTACAGCTTGAATAACCAAGTAATTTAACAGAAATC 1287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 562 aaatccggaatcccaaaatttgaacacggcggaatgggtcgctgccatcgcgcgccc 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 AAATTTGCTGATTCGACAAATTAGCGGTAGCGGATTTCACTGTTGCAATCGGTAATCCA 1347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 622 ttcggctttgaaacagcgtgacccggcgatcgctgtccgcgaag-----cagaagc 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 TTTGGTTAGGTCAAACTGTGACATCAGGTATGTTTCTGCAATGGGTCAACAGGT 1407
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 676 ctgcccacgaagactacacaccttcacaaacggcgaaagtcgaacgcgacacactggcggtattatt 1467
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 TCTGACAGTGGCACTTAGAAACTATATCAAAACCGATGAGCAGTAACCCGGGTAAAT 1467
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 736 tccgggcccgcgtgttcaacttaaaagacaggctcgctcggaataattcgcaaatatc 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 TCGGTGGAGCGTTAGTAAACTTAAATGCGCACTTATTGGAAATTAATACCGCAATATT 1527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 796 agccgacggcggaattcatgggcatctccttggccatcccgatggagcttgccatgaat 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1528 TCTCAAGCGGTGGCAATTCGAGGAATTCCTTTGGGATTCGAAGTAATCAAGCAAGCAAT 1587
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 856 gtcgccgaagcgtgaaacacacggcgaaagtcgaacgcgacacactggcggtattatt 915
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1588 TTAGTGCAACAAATTTAGAAATTTGGTCAAGTGCCTCGCGGATTCCTTGGTATTAAAGGT 1647
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 916 caggaagtactcactcgggtggcacagtcggttcggtcgtggataaagcagcgcgcatg 975
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1648 GCGCAACTCAATGCTGATTTAGCCAAAGCCTTTAAATGAAGCGCGCAACAGGCGCAATT 1707
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 976 attgccaatactctcccgagcccccagaaactgcggcgctcagcgcgcgacatc 1035
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1708 GTAAGTGAAGTTTACCAGAAATCTGCTGCTGAAAAAGCAGGACTTAAAAAGCGGCGATATT 1767
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1036 gtcctcagcctcgacgcgcggaatactgtcttcctcgcgaccttcctcgatgctgcgc 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1768 ATCAGCGCGATGACGGTCAAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAAATCGCA 1827
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1096 gccattacgccccggaagaagtcagctcgcgctatggcgcaaaagcgaa 1146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1828 ACCACTGGTGAGGCAAAAGAGATTAGCTTGACTTACTTACGGTGGGCAAA 1878
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 14, 2000, 15:42:07 ; Search time 1307.83 Seconds
(without alignments)
4027.321 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgtcaaaaataccaata.....tcagggaagtcocccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues 9077268

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :		EST:*	
1:	em_est1:	*	
2:	em_est2:	*	
3:	em_est3:	*	
4:	em_est4:	*	
5:	em_est5:	*	
6:	em_est6:	*	
7:	em_est7:	*	
8:	em_est8:	*	
9:	em_est9:	*	
10:	em_est10:	*	
11:	em_est11:	*	
12:	em_est12:	*	
13:	em_est13:	*	
14:	em_est14:	*	
15:	em_est15:	*	
16:	em_est16:	*	
17:	em_est17:	*	
18:	em_est18:	*	
19:	em_est19:	*	
20:	gb_est1:	*	
21:	gb_est2:	*	
22:	gb_est3:	*	
23:	gb_est4:	*	
24:	gb_est5:	*	
25:	gb_est6:	*	
26:	gb_est7:	*	
27:	gb_est8:	*	
28:	gb_est9:	*	
29:	gb_est10:	*	
30:	gb_est11:	*	
31:	gb_est12:	*	
32:	gb_est13:	*	
33:	gb_est14:	*	
34:	gb_est15:	*	
35:	gb_est16:	*	
36:	gb_est17:	*	
37:	gb_est18:	*	
38:	gb_est19:	*	
39:	gb_est20:	*	
40:	gb_est21:	*	
41:	gb_est22:	*	
42:	gb_est23:	*	
43:	gb_est24:	*	
44:	gb_est25:	*	
45:	gb_est26:	*	
46:	gb_est27:	*	
47:	gb_est28:	*	
48:	gb_est29:	*	
49:	gb_est30:	*	
50:	gb_est31:	*	
51:	gb_est32:	*	
52:	em_est20:	*	
53:	em_est21:	*	
54:	em_est22:	*	
55:	em_est23:	*	
56:	em_est24:	*	
57:	em_est25:	*	
58:	em_est26:	*	
59:	gb_est33:	*	
60:	gb_est34:	*	
61:	gb_est35:	*	
62:	gb_est36:	*	
63:	gb_est37:	*	
64:	gb_est38:	*	
65:	em_est27:	*	
66:	em_est28:	*	
67:	em_est29:	*	
68:	em_est30:	*	
69:	gb_est39:	*	
70:	gb_est40:	*	
71:	gb_est41:	*	
72:	gb_est42:	*	
73:	gb_est43:	*	
74:	gb_est44:	*	
75:	em_est31:	*	
76:	em_est32:	*	
77:	em_est33:	*	
78:	em_est34:	*	
79:	gb_gss1:	*	
80:	gb_gss2:	*	
81:	gb_gss3:	*	
82:	gb_gss4:	*	
83:	em_gss1:	*	
84:	em_gss2:	*	
85:	em_gss3:	*	
86:	em_gss4:	*	
87:	gb_gss5:	*	
88:	gb_gss6:	*	
89:	gb_gss7:	*	
90:	gb_gss8:	*	
91:	gb_gss9:	*	
92:	em_gss5:	*	
93:	em_gss6:	*	
94:	em_gss7:	*	
95:	em_gss8:	*	
96:	em_gss9:	*	
97:	em_gss10:	*	
98:	em_gss11:	*	
99:	gb_gss10:	*	
100:	gb_gss11:	*	
101:	em_gss12:	*	
102:	gb_gss12:	*	
103:	gb_gss13:	*	
104:	gb_gss14:	*	
105:	gb_gss15:	*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.6	7.2	490	34	AA479844
2	97.4	7.0	489	28	AA110074

AA479844 zu43h07.f
AA110074 mo51b12.f


```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 489)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE        The WashU-HMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1405092.
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:337879
Seq primer: -28M13 rev1 from Amersham
High quality sequence stop: 446.
FEATURES
source
1..489
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:557087"
/clone_lib="Life tech mouse embryo 10 5dpc 10665016"
/tissue_type="embryo"
/dev_stage="10.5dpc embryos"
/lab_host="DH10B"
/notes="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. 10.5dpc embryos, pCMV-SPORT2 vector."
BASE COUNT   130 a 116 c 140 g 103 t
ORIGIN
Query Match      7.0%; Score 97.4; DB 28; Length 489;
Best Local Similarity 53.8%; Pred. No. 1.2e-14;
Matches 231; Conservative 0; Mismatches 186; Indels 12; Gaps 1;

QY 366 attgaactcgttcggttcattcatcagcaaaacggctacatcctgaccataccca 425
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 AGTGGCCAGTGGGTCAAGGATTTCATCGTATCGGAGGATGGACTGATTGTGACAAATGCTCA 105

QY 426 cgtcgttgccggtatgggcaagtataaaagtcctgctcaacgacagcggaataaccgc 485
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 CBTGGTCAACCAAAACCGGGTCAAGGTGAGCTGAAGAAATGGAGCTACCTATGAAGC 165

QY 486 caaactcatcgttcggtatccaatccgatgcgcctctcgtctgaaatcgacgcaacgga 545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CAAATACAGGATCTGGATGAAGGCGGACATGCGCTTATCAGATTTGACCAACGAGG 225

QY 546 agagctaccgctgcataaatcggaatcccaaaatttgaaccggcggaatgggtcgc 605
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 AAAGCTGCCAGTCTCTGCTGCGCCGCTCTCTCAGAGCTGAGACCTGGAGATTTGTAGT 285

QY 606 tgccatcgccgcgccttcggtttgacacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgc 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 TGCCATTGGAAGCCCTTTTCTCTTCAAAACACAGTCAAGTGGATCGTTCAGCACACAC 345

QY 666 aggcagaagc-----ctgcccacgaaggtacacaccccttcacccaaccca 713
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 CACAGCAGGCGGCAAGAGCTGGAGCTTCGGAATCCGATATGACATACATTCAGACAGA 405

QY 714 cgttgccataaatccggggaattccggcgccgctgttcaacttaaaagacaggtgcgt 773
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 CGCTATCATCAATATGGAGATTCCGGAGGCCCGTTAGTAAACCTGGATGCCGAGGTGAT 465

```

```

QY 774 cggcatcaa 782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 TGGGATTAA 474

RESULT 3
LOCUS    A1159029      456 bp      mRNA      EST      02-OCT-1998
DEFINITION v283a08.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
IMAGE:133046 5' similar to TR:Q92743 Q92743 NOVEL SERINE PROTEASE.
; mRNA sequence.
ACCESSION A1159029
VERSION   A1159029.1 GI:3692211
KEYWORDS  EST.
SOURCE    House mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 456)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE     The WashU-HMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2286627.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:692590
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 241.
FEATURES
Location/Qualifiers
1..456
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:133046"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCGCGAATGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 124 a 107 c 127 g 97 t 1 others
ORIGIN
Query Match      6.8%; Score 95.4; DB 42; Length 456;
Best Local Similarity 53.9%; Pred. No. 3.7e-14;
Matches 226; Conservative 0; Mismatches 181; Indels 12; Gaps 1;

QY 376 ggtcgggctcatcagcaaaacggctacatcctgaccaatccacgcgtgtgcc 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 GGGTCAGGATTTCGTATCGGAGGATGGACTGATTGTGACAAATGCTCAGGTGGTACC 69

```



```

/db_xref="taxon:9606"
/clone="IMAGE:38967"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="PH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lfamid BA; Site.1: Not
I; Site.2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
ACTGGAAGATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

```

BASE COUNT      126 a 121 c 135 g 103 t
ORIGIN

Query Match      6.6%; Score 92.2; DB 22; Length 485;
Best Local Similarity 57.4%; Pred. No. 2.5e-13;
Matches 166; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 376 ggtcgggttcacatcagcaaaacggctacatcctgaccataaccacgtcgttgc 435
|| || || || || || || || || || || || || || || || || || || || ||
Db 5 GGGTCGTGGTTATTGTGTCGGAAGATGGACTGATCTGACAAATGCCACGTGTGACC 64

QY 436 ggtatgggcagatcaaaagtcctgctcaacgacgaagcgcaatataccgcgaactc 495
|| || || || || || || || || || || || || || || || || || || || ||
Db 65 AACAAACACCGGGTCAAAAGTTGAGCTGAAGAACGGTGCACATTACGAAGCCAAAATCAAG 124

QY 496 gttcggatgtccaatccgctgctgcctcttgaataacacgcaagcggaagcgtacc 555
|| || || || || || || || || || || || || || || || || || || || ||
Db 125 GATGTGGATGAGAAAGCAGACATCGCCTCATCAAAATGACCACCGGCAAGCTGCCT 184

QY 556 gtcgtcaaatcgcaatcccaaaatttgaaacggcgcaagtggtcgtgccatcgcc 615
|| || || || || || || || || || || || || || || || || || || || ||
Db 185 GTCTCTGCTTGGCGCTCCTCAGAGCTGCGCGCGGAGAGTCTGTGGTGGCCATCGGA 244

QY 616 ggcgcctcggcttgacacagcgtgacccgcgccatcgtgtccgcga 664
|| || || || || || || || || || || || || || || || || || || || ||
Db 245 AGCCCGTTTTTCCCTTCAAAACACAGTCACACCGGGGATCGTGAGCACCA 293

```

```

RESULT 6
F07134
LOCUS      F07134      350 bp      mRNA      EST      20-FEB-1995
DEFINITION HSCIWG081 normalized infant brain cDNA Homo sapiens CDNA clone
c-1wg081, mRNA sequence.
ACCESSION F07134
VERSION F07134.1 GI:672779
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 350)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Juneau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
9527534
On Sep 21, 1992 this sequence version replaced gi:279153.
Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'interationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

```

```

Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-1wg08
Insert Length: 792 Std Error: 0.00
Seq primer: (-21)M13_universal
High quality sequence stop: 371.
FEATURES             Location/Qualifiers
     source            1..350
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
     note="Organ: brain; Vector: Lfamid BA; Site.1: HindIII;
     Site.2: NotI; sex:Female; dev_stage=3 months old;
     isolate=muscular atrophy patient; tissue_type=total
     brain; total mRNA was oligo-(dT) primed and directionally
     cloned 5' -> 3' into the HindIII -> NotI sites of the
     Lfamid BA vector. Clone library from B.Souares, Psychiatr;
     Dept. Columbia University, USA. Normalization_method:
     Bento Soares, P.N.A.S in press"
BASE COUNT      88 a 95 c 100 g 66 t
ORIGIN

```

```

Query Match      6.4%; Score 89.6; DB 21; Length 350;
Best Local Similarity 56.7%; Pred. No. 9.8e-13;
Matches 164; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 376 ggtcgggttcacatcagcaaaacggctacatcctgaccataaccacgtcgttgc 435
|| || || || || || || || || || || || || || || || || || || || ||
Db 6 GGGTCTGGTTATTGTGTCGGAAGATGGTCTGATCTGACAAATGCCACGTGTGACC 65

QY 436 ggtatgggcagatcaaaagtcctgctcaacgacgaagcgcaatataaccgcaactc 495
|| || || || || || || || || || || || || || || || || || || || ||
Db 66 AACAAACACCGGGTCAAAAGTTGAGCTGAAGAACGGTGCCTTACGAAGCCAAAATCAAG 125

QY 496 gttcggatgtccaatccgctgctgcctcttgaataacacgcaagcggaagcgtacc 555
|| || || || || || || || || || || || || || || || || || || || ||
Db 126 GATGTGGATGAGAAAGCAGACATCGCCTCATCAAAATGACCACCGGCAAGCTGCCT 185

QY 556 gtcgtcaaatcgcaatcccaaaatttgaaacggcgcaagtggtcgtgccatcgcc 615
|| || || || || || || || || || || || || || || || || || || || ||
Db 186 GTCTCTGCTTGGCGCTCCTCAGAGCTGCGCGCGGAGAGTCTGTGTGCGCATCGGA 245

QY 616 ggcgcctcggcttgacacagcgtgacccgcgccatcgtgtccgcga 664
|| || || || || || || || || || || || || || || || || || || || ||
Db 246 AGCCCGTTTTTCCCTTCAAAACACAGTCACACCGGGGATCGTGAGCACCA 294

```

```

RESULT 7
AI903651/c
LOCUS      AI903651      337 bp      mRNA      EST      01-DEC-1999
DEFINITION QY-BT032-190299-136 BT032 Homo sapiens cDNA, mRNA sequence.
ACCESSION AI903651
VERSION AI903651.1 GI:6494038
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1134214.
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```


ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 288)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert size: 1530

High quality sequence steps: 143 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Insert length: 1530 Std Error: 0.00
Seq primer: M13RP1

High quality sequence stop: 143.

FEATURES
source
1..288
/organism="Homo sapiens"
/db_xref="GDB:397142"
/db_xref="taxon:9606"
/clone="IMAGE:24795"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5';
AACTGGAAGAAATTCGCGCGCAGAGAAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
68 a 71 c 75 g 61 t 13 others

BASE COUNT
ORIGIN

Query Match 5.8%; Score 81.2; DB 21; Length 288;
Best Local Similarity 53.7%; Pred. No.1.2e-10;
Matches 152; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 374 tcggttcgggcttcacatcagcaaaaacggctacatcctcaccacatccacacgtcgttg 433
|||||
DB 5 TCGGTCCTGGGTTATTGTGTCGGAGATGGACTGATCGTGACAAATGCCACGTTGGTA 64
|||||

QY 434 ccggtatggggcagttacaaagtctcgtctcaacagcaagcgcaatataccgccaaactca 493
|||||

DB 65 CCAACAAGCACCCGGGTCAAAAGTTGAGCTTGAGATGAGAAGCGGTGCCACTTACGAAGCAAAATCA 124
|||||

QY 494 tcggttcgagatgcacaatcccgatgcgcctctcaaaatcacgcaacgaagagctac 553
|||||

DB 125 AGGATGTGATGAGAAAGCAGACATCGCACTCATCAAAATTGCCACCGGCGCAAGTCNC 184
|||||

QY 554 ccgctgcgtcaaaatcggcaatcccaaaaatttgaaccggcgcaatgggctcgctgcactcg 613
|||||

DB 185 CTGCTCTGCTNCTTTGCCNCTCCTCAGAGCTCGGGNCNGGAGAGTTCGTNGTCNCACGCG 244
|||||

QY 614 gcgcgccttcggtcttgacaaacagcgtgaccgcgcgcatcgt 656
|||||

DB 245 GANGCCNTTTTCCCTTTCAATACACAGTWNANCNCCGGGNTCG 287
|||||


```

AI982113      AI982113      623 bp      mRNA      EST      15-SEP-1999
LOCUS         pat.pk0074.e8.f chicken activated T cell cDNA Gallus gallus cDNA
DEFINITION   clone pat.pk0074.e8.f 5' similar to serine protease, mRNA sequence.
ACCESSION    AI982113
VERSION      AI982113.1 GI:5885141
KEYWORDS     EST.
SOURCE       chicken.
ORGANISM     Gallus gallus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
AUTHORS     Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
TITLE        1 (bases 1 to 623)
JOURNAL      Tirunegard,V.G., Sofer,L. and Burnside,J.
COMMENT      An expressed-sequence-tag database of activated chicken T cells:
              sequence analysis of 5596 clones
              Unpublished (1999)
              On May 20, 1999 this sequence version replaced gi:4878134.
              Contact: Joan Burnside
              Molecular Endocrinology
              University of Delaware
              40 Townsend Hall, Newark, DE 19717, USA
              Tel: 302 831-1345
              Fax: 302 831-3651
              Email: joan@udel.edu
              Seq primer: 27.
FEATURES     Location/Qualifiers
              1..623
              /organism="Gallus gallus"
              /db_xref="taxon:9031"
              /clone="pat.pk0074.e8.f"
              /clone_lib="chicken activated T cell cDNA"
              /sex="male"
              /cell_type="Con A-activated splenic T cell"
              /lab_host="E.coli TOP10 F'"
              /note="Vector: pcDNA3"
BASE COUNT   187 a 137 c 156 g 131 t 12 others
ORIGIN
Query Match      5.6%; Score 77.8; DB 63; Length 623;
Best Local Similarity 54.3%; Pred. No. 1.2e-09;
Matches 157; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 376 ggttcgggttcattcgcgcaaaacggctacatcctgacccaataccacgctgtgccc 435
Db 67 GGTTCGGGTTCTTCGCGAGGATGGCTGATAGTACCAACGCTCACCCTGGTCAACC 126
QY 436 ggtatgggcagtatcaaatcctctgctcaacgacgaagcgaataataccgccaactc 495
Db 127 ACAAGAACAGGGTGAAGGTGGAGCTGAAGATGGAGAACATACCAAGCTAAATATAA 186
QY 496 ggttcgatgtccaatccgatgctgcgcctctgaaatcgacgcaacggaagagctacc 555
Db 187 GAGCTTGATGAGAAAGCTGACATGCTGACATAATTAATAGATGCTCAGGGTAAATGCCA 246
QY 556 gtctgcaaatcggcaatcccccaaaattgaaacccggcggaatgggtcgtcgtccgcgc 615
Db 247 GTCTTCTGCTTGTGCTGAGTCTGAGAGACTTGAGCCGCGGAGAGTTTGTGTTGCTATGGG 306
QY 616 ggcgccttcggctttgacacagcgtgacccgagcgtcgtgctccgcca 664
Db 307 AATCCATTTTCCCTCCAAAACACAGTGAACACGGGGATCGTCAACACCA 355

RESULT 14
AA349615
LOCUS         AA349615      292 bp      mRNA      EST      21-APR-1997
DEFINITION   EST36620 Infant brain Homo sapiens cDNA 5' end similar to similar
              to Protease DO, mRNA sequence.
ACCESSION    AA349615
VERSION      AA349615.1 GI:2002068
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 292)
JOURNAL      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
              Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
              White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
              Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
              Fitzgeraid,L.M., Fitzhugh,W.M., Hanna,M.C., Fritchman,J.L., Geoghagen,N.S.,
              Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
              Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
              Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
              Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
              Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
              Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
              Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
              He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
              Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
              Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
              Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
              Fraser,C.M. and Venter,J.C.
              Initial assessment of human gene diversity and expression patterns
              based upon 83 million nucleotides of cDNA sequence
              Nature 377 (6547 Suppl), 3-174 (1995)
              12140200
              On Sep 12, 1996 this sequence version replaced gi:1394192.
              Other_ESTs: THCI72755
              Contact: Kerlavage, AR
              Bioinformatics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850 USA
              Tel: 3018699056
              Fax: 3018699423
              Email: arkerlavet@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST, please check the TIGR Human Gene
              Index (http://www.tigr.org/tdb/hgi/hgi.html)
              Seq primer: M13 Reverse.
FEATURES     Location/Qualifiers
              1..292
              /organism="Homo sapiens"
              /db_xref="ATCC (inhost):150705"
              /db_xref="taxon:9606"
              /clone_lib="Infant brain"
              /sex="female"
              /dev_stage="infant"
              /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
              Site_2: NotI"
BASE COUNT   76 a 88 c 79 g 47 t 2 others
ORIGIN
Query Match      5.0%; Score 69.6; DB 32; Length 292;
Best Local Similarity 57.0%; Pred. No. 1.1e-07;
Matches 126; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 444 cagttcaaatcctcgtctcaacgacgaacgacgaataacgcgaactcattcgttcgga 503
Db 1 CCGGGTCAAAAGTTAGCTGAAGAACGGTGCCTTACGAAGCCAAATCAAGGATGTGA 60
QY 504 tgtccaatccgatgtgcgcctctgaaatcgacgcaacggaagagactaccgtctcaca 563
Db 61 TGAGAAAGCAGACATCGCACATCAATCAAAATTGACCACCGGGCAAGTGCCTCTCTGCT 120
QY 564 aatcgggcaatcccaaaatttgaaacccggcggaatgggtcgtcgtcgtcgtcgtcgt 623
Db 121 GCTTGCCCTCTCTCAGAGCTCGGGCCGGGAGAGNTCGTGGTCCCATCGGAAGCCGTT 180
QY 624 cggctttgacacagcgtgaccgcggcgcgtcgtgtccgcca 664
Db 181 TTCCCTTCAAAACACAGTCACCAACCGGGATCGTGAGCACCA 221

```


GOGDH
LOCUS GOGDH 2890 bp DNA BCT 15-MAR-1999
DEFINITION G. oxydans gdh gene.
ACCESSION X62710 S60040
VERSION X62710.1 GI:58416
KEYWORDS coenzyme PQQ; gdh gene; gluconic acid production; glucose dehydrogenase; quinoprotein.
SOURCE Gluconobacter oxydans.
ORGANISM Gluconobacter oxydans
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae; Gluconobacter.
REFERENCE 1 (bases 1 to 2890)
AUTHORS Cleton-Jansen, A.M., Dekker, S., van de Putte, P. and Goosen, N.
TITLE A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in Gluconobacter oxydans
JOURNAL Mol. Gen. Genet. 229 (2), 206-212 (1991)
MEDLINE 92017653
REFERENCE 2 (bases 1 to 2890)
AUTHORS Goosen, N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1992) N. Goosen, Dept of Mol Genetics, Gorlaeus Laboratories, Leiden University, P O Box 9502, 2300 RA Leiden, THE NETHERLANDS

FEATURES
source Location/Qualifiers
1. .2890
/organism="Gluconobacter oxydans"
/db_xref="taxon:442"
RBS 242. .245
gene 251. .2677
/gene="gdh"
CDS 251. .2677
/gene="gdh"
/EC_number="1.1.99.17"
/codon_start=1
/transl_table=11
/product="glucose dehydrogenase (pyrroloquinoline-quinone)"
/protein_id="CAA44594.1"
/db_xref="GI:4493391"
/db_xref="SWISS-PROT:P27175"
/translation="MSTISRPRLLWALITAAVFALCGAILTVGGAWVAAIGGPLYYVIL
GLALLATAFLSFRNPAALYLFVAVVFGTVIWEITVVGDLIWLIPRSDIVIIILGIWL
LLPFVSRARSVARGRPSRSPAVGVAVLALFASLFTDPHDISGELPTQIANASPADPD
NVPASEWHAYGRTQAGDRWSPLNQINATNVSNLKVAWHIHTKMMNSNDPGATNEAT
PIEFNNTLYMCSLHOKLEFVDGATGNVKNVYDPKLQINPGFQHLTCRGVSFHETPANA
MDSGDNPAPTDCAKDSILPVNDGRLVEVDADTGKTCGFGNGEIDLRVPNQPYTTTPG
QYEPTSPPVITDKLIIANSAITDNGSVKQASGATQAFDVTGKRVVVFDAENPDNPQL
PDESHPVFHPNSPNSWIVSSYDANLNLVYIPMGVGTDPQWGGDRTKDSEFAPGIVAL
NADTGKLAWFYQTVHDLWDMELPSQPSLVDVTQKDGTLVPAIYAPTGTGDI FVLDRR
TGKEIVPAPEPTVPQGAAPGDHTSPTQPMSQLTLRPKNPLNDSI WGGTIFDQMFCSI
YFHTLRLEGPFPTPSLKGSLIFPGDLGMFEWGGLAVDPQROVAFANPISLFFVSQVLP
RGPGNLWPEENAKGTGGETGLQHNHYGIPYAVNLHPFLDPVLLPFGIKMPCRTPPWGY
VAGIDLKINKVWQHNRNGLRDSMYGSSLPILPPIKIGVPSLGGPLSTAGNLGFLTA
SMDYYIRAYNLTKGVWLQDRLPAGAQTPTITYAINGKQIVTYAGGHNSFPTRMGDD
IIAYALPDQK"
misc_feature 293. .664
/gene="gdh"
/note="membrane spanning fragment"
mutation 2612
/gene="gdh"
/note="mutation changing substrate specificity"
BASE COUNT 555 a 975 c 786 g 574 t
ORIGIN

Query Match 1.4%; Score 20; DB 1; Length 2890;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 597 atgggtcgctgccatcggcg 616
|||||
Db 340 ATGGGTGCTGCCATCGGC 359

10

11

12

13

14

15

16

2353

5.5

91

1-2-17